SCIENTIFIC AMERICAN

AUGUST 1992

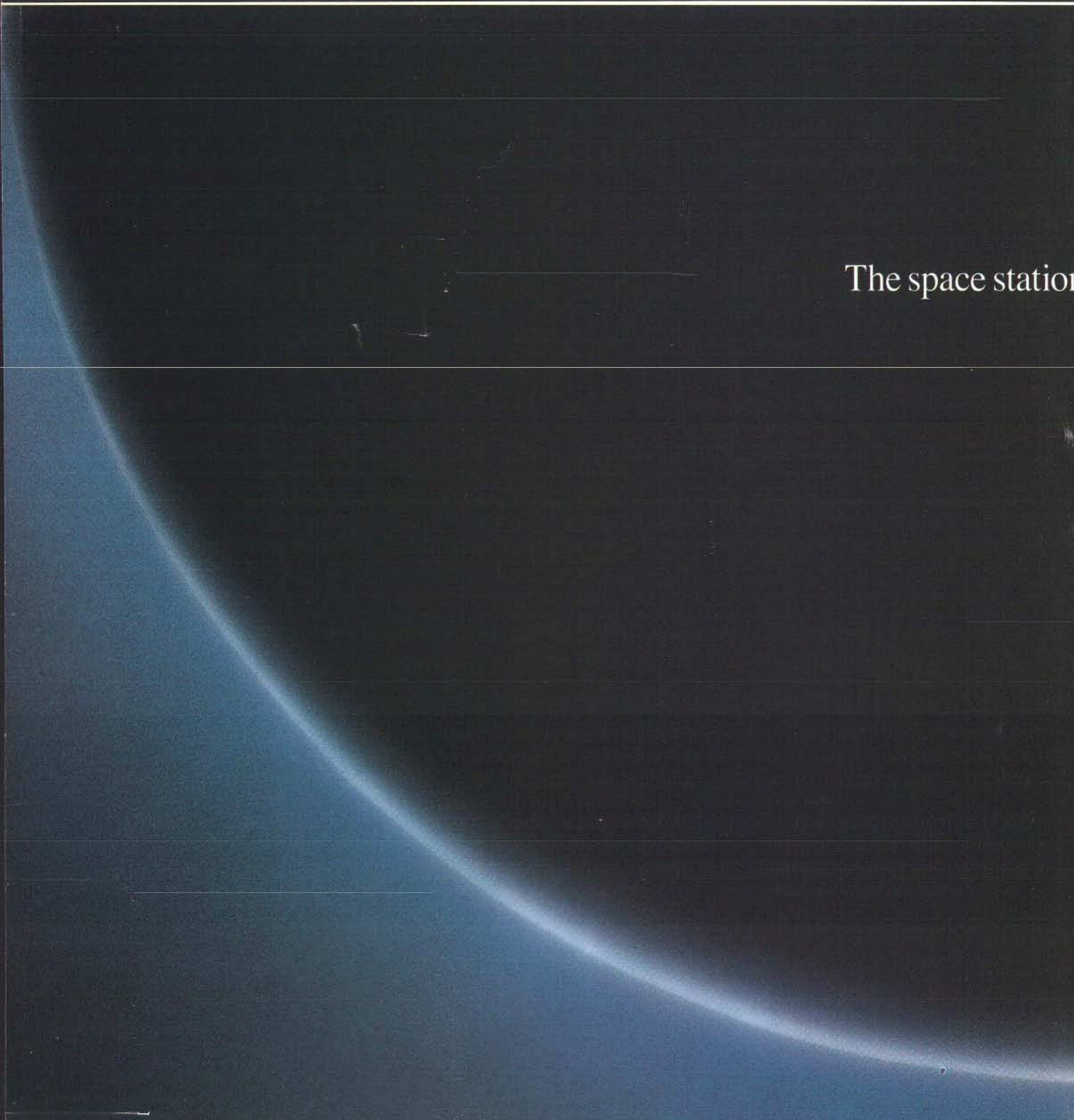
$3.95

Will an American maglev finally fly?

Musical illusions: how pitch tricks the ear.

Encrypted ID's for digital privacy.

Restless Kilauea: *understanding its dynamic processes helps to predict other volcanic eruptions.*



The good health of the home planet is the reason we're building the manned space station today—and why we'll launch the first components for it in less than four years.

Some of the laboratories will be dedicated to the life sciences—studies that touch the lives of virtually everyone on earth. Others are designed to learn more about materials—the basic building blocks of our civilization.

The space station will have unique micro-gravity labs where scientists can do experiments that are impossible on earth.

And it will allow research into the natural processes that created and still change our world. And our lives.

Best of all, research on board the space station will increase our rate of discovery and invention in materials and life sciences.

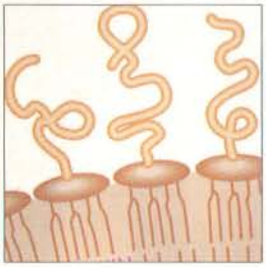
SCIENTIFIC AMERICAN

August 1992 Volume 267 Number 2

18 

Dynamics of Kilauea Volcano
John J. Dvorak, Carl Johnson and Robert I. Tilling

One of the longest volcanic eruptions in recorded history began in 1983. Lava flows from Kilauea have since added 120 hectares of new land to the island of Hawaii and covered 100 square kilometers. From a nearby cliff, the authors observed and analyzed these events. Their findings clarify the mechanisms of volcanism. The techniques they developed are helping to predict other eruptions.

26 

Bacterial Endotoxins
Ernst Theodor Rietschel and Helmut Brade

Bacterial endotoxins are a two-edged sword. These cell-wall components of a major group of bacteria account for many symptoms of cholera, whooping cough and other diseases. But they can also enhance the immune response to other bacteria, viruses and even cancer. Recent findings may lead to ways of curbing the harmful effects of endotoxins and harnessing their disease-fighting capacity.

34 

How Cosmology Became a Science
Stephen G. Brush

The big bang became the established explanation for the origin of the universe almost overnight, when Arno A. Penzias and Robert W. Wilson observed the faint signals of the cosmic background radiation in the 1960s. Their achievement relied on a rich legacy of theory and experiment that enabled big bangers to challenge successfully the earlier concept of a universe that had always existed.

42 

Naked Mole Rats
Paul W. Sherman, Jennifer U. M. Jarvis and Stanton H. Braude

These African rodents have been described as saber-toothed sausages, as baby walruses or simply as ugly. Naked mole rats are also fascinating creatures. Unlike most mammals, they practice the "eusocial" behavior typically observed in ants and termites. In mole rat burrows, only a few individuals breed; others care for the offspring. What are the genetic and evolutionary roots of this social organization?

62 

SCIENCE IN PICTURES

Evolution Comes to Life
Ian Tattersall

The curator of the American Museum of Natural History's new Hall of Human Biology and Evolution describes the daunting task of constructing lifelike figures of our distant ancestors, guided only by fragments of ancient bone.

Scientific American (ISSN 0036-8733), published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017. Second-class postage paid at New York, N.Y., and at additional mailing offices. Authorized as second-class mail by the Post Office Department, Ottawa, Canada, and for payment of postage in cash. Canadian GST No. R 127387652. Subscription rates: one year $36 (outside U.S. and possessions add $11 per year for postage). Subscription inquiries: U.S. and Canada 800-333-1199; other 515-247-7631. Postmaster: Send address changes to Scientific American, Box 3187, Harlan, Iowa 51537.

70 

Paradoxes of Musical Pitch
Diana Deutsch

Just as optical illusions can trick the eye, so various combinations of musical pitch can deceive the ear. Recent research shows that these auditory paradoxes may be related to the brain's processing of speech. The way individuals hear various sequences of tones seems peculiar to their particular language or dialect.

76 

Achieving Electronic Privacy
David Chaum

All of your electronic transactions, from credit card purchases to bank withdrawals, are creating a digital dossier of your life. The author proposes an encryption system that would allow individuals and institutions to take advantage of the benefits of computer communications while protecting privacy.

84 

TRENDS IN TRANSPORTATION

Air Trains
Gary Stix, staff writer

In the 1960s, a handful of engineers devised what they believe is the answer to ground transportation needs for the next century: high-speed trains buoyed on magnetic fields. While Europe and Japan forged ahead, Washington zeroed out U.S. programs in the 1970s. Now, with the backing of an equally zealous senator, these aging visionaries may have a second chance to see whether maglev will fly.

DEPARTMENTS

8 

Science and the Citizen

Why fetal cell research should proceed.... Ambiguous neutrinos.... Real (good) grub.... Still testing nuclear weapons.... Glimpsing the surface of Pluto.... PROFILE: Partners in immunology: Philippa Marrack and John Kappler.

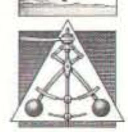
94 

Science and Business

Engineered tissues.... Tilting toward industrial policy.... The ultimate jogger.... A test for Alzheimer's disease.... The silicon retina goes commercial.... THE ANALYTICAL ECONOMIST: Market insights from experimental economics.

5 

Letters to the Editor
Controversial cauldron.... It's Charlotte, for short.

6 

50 and 100 Years Ago
1892: A peculiar problem in France.... First metal detector.

101 

Mathematical Recreations
Getting the jump on Olympic contenders from other planets.

104 

Book Reviews
Microstructures.... American rails.... The life in the rocks.

108 

Essay: *Lawrence Summers*
An economist calculates the profits in educating women.

THE COVER photograph captures a fountain of lava that spurted from the east rift zone of Kilauea volcano. Kilauea's present eruption, now in its 10th year, is proving a boon for scientists seeking to understand and to predict outbreaks of volcanism [see "Dynamics of Kilauea Volcano," by John J. Dvorak, Carl Johnson and Robert I. Tilling, page 18]. The volcano's relentless activity also serves as a sobering reminder of the potent forces lurking inside the earth; the recent lava flows have buried a region the area of Manhattan Island.

THE ILLUSTRATIONS

Cover photograph by Lee Allen Thomas/Doug Peebles Photography

Page	Source	Page	Source
18–19	Lee Allen Thomas/Doug Peebles Photography	47	Jennifer U. M. Jarvis
20	Guilbert Gates	48	Raymond A. Mendez
21	Landsat/EROS Data Center	62–69	Jason Goltz
22–23	Guilbert Gates	71	© 1960 M. C. Escher/Cordon Art
24	J. D. Griggs, U.S. Geological Survey (*top*), Guilbert Gates (*bottom*)	72–73	Edward Bell
25	J. D. Griggs (*left*), M. Timothy O'Keefer/Bruce Coleman, Inc. (*right*)	74	Zimmerman/FPG International (*left*), Edward Bell (*right*)
27	Helmut Brade	75	Edward Bell (*left*), Capitol Records, Inc. (*right*)
28	Ian Worpole (*top*), Günther Schmidt, Borstel Research Institute (*bottom*)	77–80	Jared Schneidman
29	Ian Worpole (*left*), Manfred Kastowsky and Harald Labischinski (*right*)	82	Marc Skinner
30–32	Ian Worpole	84–85	Tomo Narashima
34–35	Jason Goltz	86	Robert Prochnow
36	Courtesy of Ralph A. Alpher	87	F. Lee Corkran/Sygma
37	Johnny Johnson	88	Robert Prochnow (*left* and *right*)
38	Bettmann Archive	89	George V. Kelvin (*top*), Japanese Railway Technical Research Institute (*top left*), Transrapid International (*bottom left*), SRI International (*right*)
39	Johnny Johnson	90–91	Japanese Railway Technical Research Institute (*top left*), Transrapid International (*middle left*), George V. Kelvin (*bottom left* and *right*)
40	Cambridge University Press © 1965	92	French National Railways
42–43	Raymond A. Mendez/Work As Play	102	Jared Schneidman
44–45	Patricia J. Wynne	103	Johnny Johnson
46	Raymond A. Mendez		

SCIENTIFIC AMERICAN®

Established 1845

SCIENTIFIC AMERICAN, INC.
415 Madison Avenue
New York, NY 10017
(212) 754-0550

LETTERS TO THE EDITORS

Counting Your Change

Timothy Taylor's article on the origin of "The Gundestrup Cauldron" [SCIENTIFIC AMERICAN, March] was a fascinating exercise in antiquarian detective work. Yet deducing the cauldron's country of origin through the decomposition of its weight into an integral number of coins available only from Persia is an unjustifiable leap of faith. Although the bowl could have been melted from precisely 1,666 sigloi of 5.67 grams each, it is unlikely the coins had the same mass to the nearest hundredth of a gram. (U.S. quarters vary by approximately five hundredths of a gram.) The variation in weight creates an uncertainty of as much as three coins. Different numbers of coins from other countries would have fit the hypothesis equally well.

GREG BLONDER
Director, Materials and Technology
Integration Research Laboratory
AT&T Bell Laboratories

I am sure the error [from uncertainty about the coin masses] is dwarfed by the uncertainty in estimating the weight of the missing parts of the cauldron or indeed of the wide range of possible losses during its fabrication. No production process of the period would directly yield the sharp cornered panels used for the sides of the cauldron; rather a rounded shape would be trimmed to those finished forms. The resulting loss of material would be impossible to estimate to the accuracy needed to support the assertion that the siglos was the base material used.

STEVE THOMPSON
London

Taylor replies:

In antiquity, precious metal vessels were often made in round and easily memorable multiples of coin weights so that they might serve in effect as large denomination bank notes. Their production did not require the reworking of real coins. A sheet-silver vessel would have been made to a siglos standard from an appropriately sized ingot and trimmed to reach its exact value. Despite the apparent low accuracy of the calculation, the probability of obtaining by chance two siglos-based round figures for the Gundestrup cauldron (one and two third thousands and three quarters of a thousand) is low. The calculations should be seen as only one part of the argument that the cauldron belongs to a silversmithing tradition strongly influenced by Persia.

Her Rightful Name

By her own account in *Charlotte's Web,* by E. B. White, Charlotte's species was *Aranea cavaticus,* and not *A. diadematus* as mentioned by Fritz Vollrath in "Spider Webs and Silks" [SCIENTIFIC AMERICAN, March]. At least that is how she introduced herself, exercising license in feminizing the specific name:

"My name," said the spider, "is Charlotte."

"Charlotte what?" asked Wilbur, eagerly.

"Charlotte A. Cavatica. But just call me Charlotte."

As a matter of literary, if not scientific, propriety, shouldn't one of the noblest and most attractive heroines in all literature be allowed her own identity? I can imagine how the error occurred: the range of *A. diadematus,* the cross spider, is cosmopolitan; *A. cavaticus,* the barn spider, is limited to eastern North America and may be unfamiliar to Professor Vollrath.

ELIOT PORTER
St. Louis Post-Dispatch
St. Louis, Mo.

Warming Up the Market

Richard Elliot Benedick's essay on global warming, "A Case of Déjà Vu" [SCIENTIFIC AMERICAN, April], is worthy and telling. To me, the uncertainty about global warming is nothing but academic. In the 1800s and up to 1900, the Columbia and Willamette rivers, among others, regularly froze in winter, allowing easy sled and wagon travel. As described in exploration reports and mountaineering literature, the extent of glaciers and snowpack in the Northwest and Rockies is minuscule compared with their former glories.

The market system is far more resilient and capable than we give it credit, and there is no question that a vigorous response would follow incentives to reduce global warming. As the world's largest consumer, polluter and market economy, the U.S. should rightly lead the movement for change. We should feel an obligation to err on the side of caution because as consumers, we are the major "beneficiaries" of chlorofluorocarbon pollution and global warming, yet the negative effects go far beyond our own borders.

WALTER T. HASWELL III
Corvallis, Ore.

A Crick in His Thinking

It seems to me that Francis Crick's deterministic assessment of the nature of consciousness, as reported in John Horgan's profile of him ["Science and the Citizen," SCIENTIFIC AMERICAN, February], is as metaphysical as any characterization in terms of free will. He asserts that "you, your joys and your sorrows, your memories and ambitions, your sense of personal identity and free will, are in fact no more than the behavior of a vast assembly of nerve cells and their associated molecules."

If we carry this thesis to its logical conclusion, then we must also declare reason, judgment, logic, consistency and similar faculties as equally illusory. Science thus becomes a sham process. Paradoxically, if Dr. Crick is correct, his suppositions lose their validity because they are no more the result of reason than is the plodding, mindless orbit of the earth around the sun.

MARK ADKINS
Phoenix, Ariz.

ERRATA

In the box on page 69 of "Heisenberg, Uncertainty and the Quantum Revolution" [May], a mathematical formula is incorrect. It should read: $\Delta x = \frac{\lambda}{2 \sin \Theta}$.

The caption of the cover illustration for the May issue neglected to credit David Malin of the Anglo-Australian Observatory for the photograph on which the painting is based.

The article on "The Codex Mendoza" [June] misstated the relative size of Tenochtitlán, the capital of the Aztec Empire: it actually contained about 1 percent of the population then living in central Mexico.

50 AND 100 YEARS AGO

AUGUST 1942

"Scientific American, since last autumn, has quietly organized some of its thousands of amateur telescope makers. A limited number of these advanced workers have already demonstrated that the amateur who works mainly for the love of the work, or for the hate of Hitler's regime, can successfully do precision optical work of a grade of difficulty that is looked upon even by most professionals as ultra, and which some of them find it prudent not to tackle."

"I have no quarrel with—in fact, I commend—those individuals who want to store their cars for patriotic reasons. Nor will I argue about possible savings in gasoline and oil if cars are taken off the road. However, preservation is one thing and waste is another. Persons who are considering storing their cars for the duration to save tires should consider the fact that tires deteriorate when not in use. This is because normal operation of a car flexes the rubber and keeps it alive. Light—even artificial light—damages stored tires. All windows in the storage place should be blacked out and the car should be blocked up to remove the weight from the tires."

"It is evident that the Crab Nebula is expanding in all directions, and at a very rapid rate, astronomically speaking. Carrying the motion back in time as well as in space, it appears that all parts of the nebula started from a central point at nearly the same time, about 800 years ago. Chinese records described the appearance of a 'guest-star' in the right part of Taurus, and in the year 1054 A.D. This is almost 900 years ago, but it would hardly strain the observations of the fuzzy moving nebular condensations to fit them to this slower rate of motion."

"In a paper reported in *The Journal of the Institution of Engineers*, Australia, Alan Price states that it is a mistaken idea that road corrugation results from the frictional action of motor car wheels. He points out that, if this were so, the corrugations would occur only adjacent to the wheel tracks, whereas actually they are found to extend right across the road width, and he postulates the theory that they are the direct result of air currents produced by the motion of the car. Observations and experiments are quoted by Mr. Price to show that air currents have wavelengths which eventually bring about 'heaps' or corrugations at right angles to the directions of flow; in the case of a road, however, these corrugations do not occur exactly normal to the center line, but lie slightly forward from left to right. The reason for this latter phenomenon is that the air stream is not running parallel to the center line, because the camber of the road induces the current to follow the line of least resistance."

Electrical ore finder

AUGUST 1892

"French science has to deal with a peculiar problem, how to prevent the depopulation of the country, which is now going on so rapidly that deaths exceed the births by nearly 40,000 in a single year. Increasing the birth rate having proved impracticable, the present hope is to diminish the death rate. At a recent meeting of the new Society for the Protection of Children, Dr. Rochard stated that 250,000 infants die yearly, of whom at least 100,000 could be saved by intelligent care. Stringent laws have been already passed to aid in preventing this great waste of life. It is now illegal for any person to give children under one year of age any solid food except on medical advice, and nurses are forbidden to use nursing bottles having rubber tubes. Efforts are being made also to induce Parisian mothers to nurse their own infants."

"In the figure (*left*) is shown an instrument for indicating the presence of metals. The induction coil consists of a primary coil made of coarse wire and connected with a rapid automatic circuit breaker and battery. The secondary coil is made of fine wire and is arranged exactly at right angles to the coarse wire coil. A telephone is connected with the secondary coil. If the primary circuit is continuously and rapidly interrupted while the coil is not in the vicinity of any magnetic material, no sound will be heard in the telephone, as all the inductive influences are equal and opposite; but when the coil is held in proximity to metal or magnetic ore, the equilibrium is disturbed and the sound is heard in the telephone."

"It is said that there are at least two distinct races of stingless bees in South America, but these races have not much value as honey gatherers, and moreover they build combs with very thick-walled cells, and probably they would not be worth cultivating as compared with the European, Asiatic, and African races. But if we can cross our bees with the giant bees of India and obtain a race with a long proboscis and perhaps increased size (if that should prove to be of any advantage), and cross this improved race with the South American stingless bees, we shall then have a race of bees which it will be difficult to improve."

"The Empire State express breaks its own record so often that close attention is required to keep track of it, so says the *Railroad Gazette*. On July 4 it ran from Syracuse to Rochester, 81 miles, in 74 minutes, equal to $65\frac{2}{3}$ miles an hour."

Now schools without French teachers can still have lessons in French.

Students at three high schools in northwestern suburbs of Chicago now can take classes their schools might not have been able to afford otherwise — thanks to a system provided by Ameritech.

Called Distance Learning, the system is a new application for video conferencing, a service originally intended for business use. By video, a teacher can combine students from three schools to form a class and teach them simultaneously even though the schools are twenty miles apart. It's only one of many educational efforts Ameritech has under way. Our Buddy System, for example, is linking students and home computers with their schools and a whole world of information. We're also providing grants to students doing graduate work in telecommunications and supporting teachers writing curricula using our Ameritech PagesPlus® directories as learning tools.

At Ameritech, we're developing innovative products and services of all kinds to help make sure today's students aren't just ready to face the future, but to shape it.

For more information on our contributions to the quality of life in communities we serve, call 1-800/242-8580.

THE COMMUNICATIONS COMPANIES OF AMERITECH ARE:
Illinois Bell • Indiana Bell • Michigan Bell • Ohio Bell • Wisconsin Bell Ameritech Audiotex Services • Ameritech Credit • Ameritech Development Ameritech Information Systems • Ameritech International Ameritech Mobile Communications • Ameritech Publishing Ameritech Services • The Tigon Corporation.

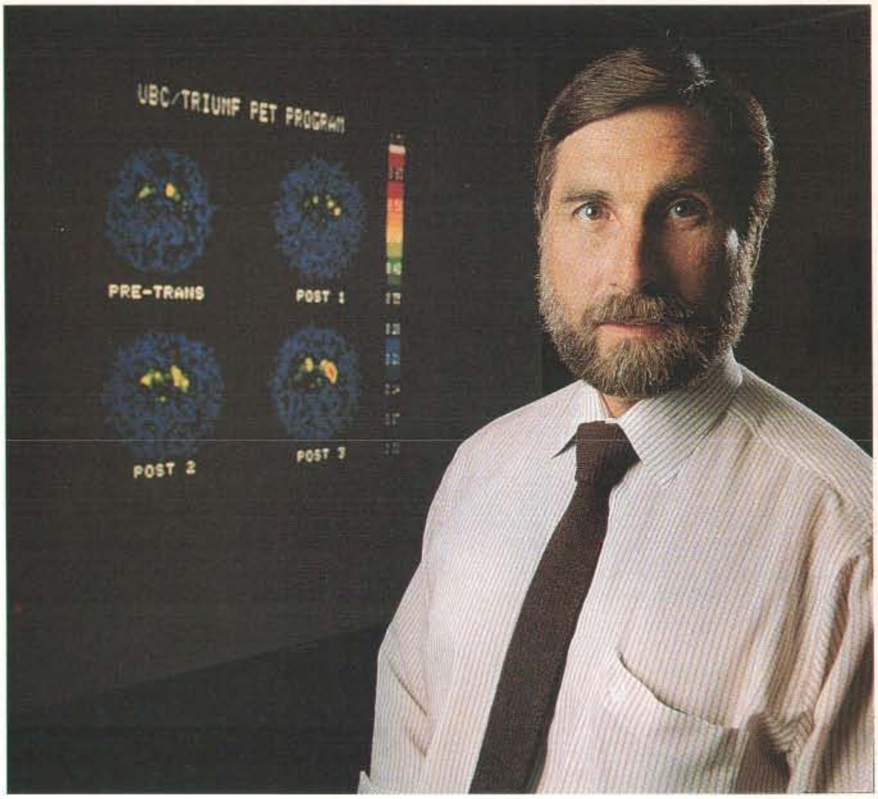
AMERITECH

Connecting You With Tomorrow

SCIENCE AND THE CITIZEN

Aborting Research

Fetal cell transplants in the U.S. are stalled by opposition



J. WILLIAM LANGSTON of the California Parkinson's Foundation uses positron emission tomography to monitor fetal tissue grafts. Photo: Douglas L. Peck.

At the First International Congress of the Cell Transplant Society held recently in Pittsburgh, 13 papers were delivered during a special session on transplanting tissue obtained from fetuses. Their results inspired increasing confidence that those vigorous cells can enhance life for people with a number of disabling diseases. But only one presentation was by U.S. researchers—and it was based on sheep. The balance were by workers from China, Japan, Russia, Hungary, Singapore, Sweden and Yugoslavia.

The paucity of reports by U.S. scientists was mute testimony to the effects of a 1988 federal moratorium on funding for studies on fetal cell transplantation that use tissue obtained from voluntarily aborted fetuses. The Bush administration has continued to enforce the ban even though a specially convened expert panel recommended that the research should proceed provided safeguards were put in place to prevent it from encouraging abortion. In addition, the ban has discouraged private funding for the work and had a chilling effect on most investigative efforts into fetal development, says D. Eugene Redmond, Jr., a fetal transplantation researcher at Yale University.

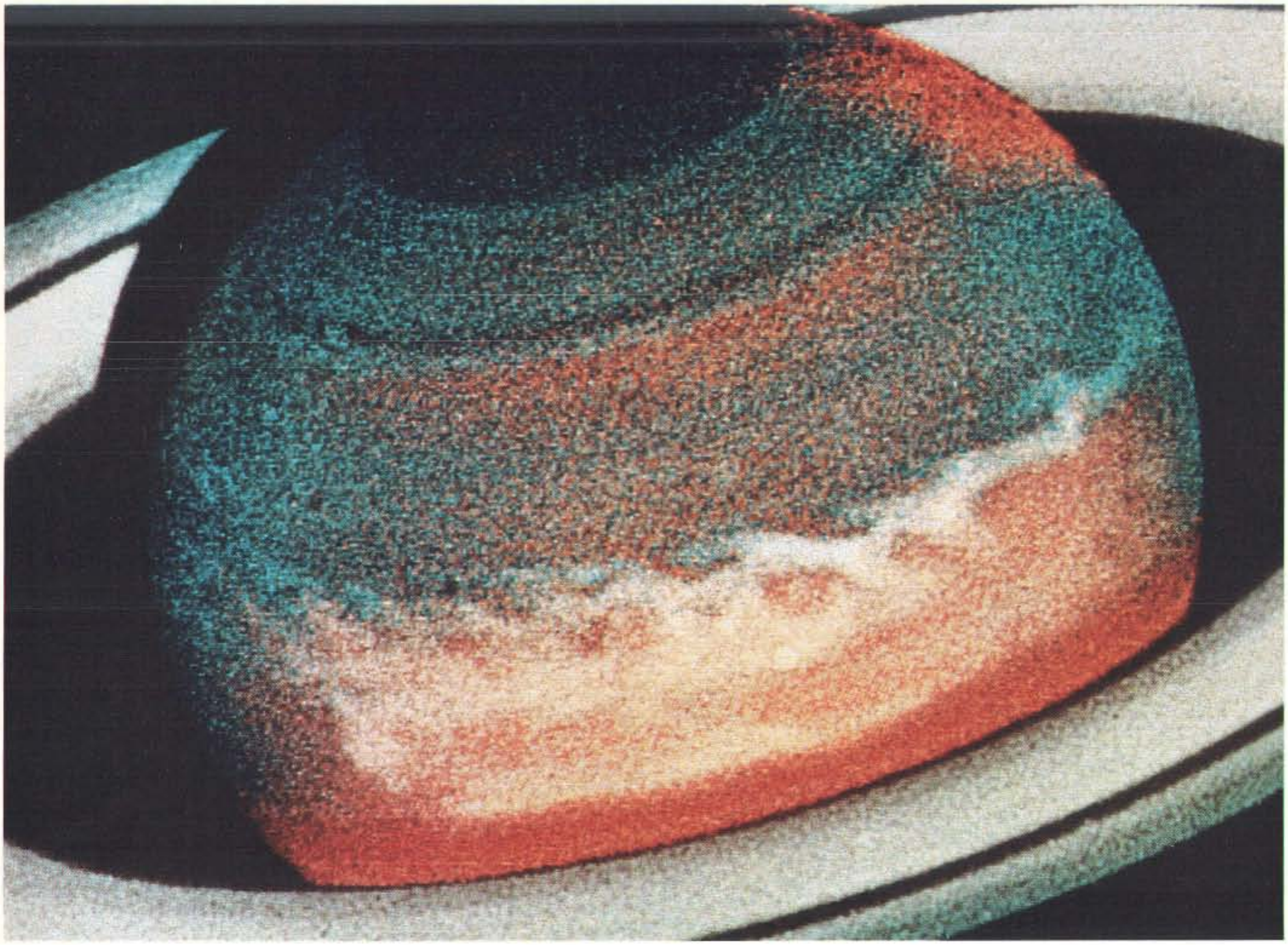
The standoff has now escalated into a full-blown political confrontation. President Bush vetoed a congressional effort to overturn the ban in late June, and an attempt to override the veto failed to succeed. Opponents of the administration's position—among them the American Medical Association, the Association of American Medical Colleges, the Association of Biotechnology Companies, the Juvenile Diabetes Foundation and several foundations on Parkinson's disease—emphasize that their support for research on fetal cell transplantation does not imply any particular view on abortion. "What we are talking about has nothing to do with abortion," insists Sara King, research director of the Juvenile Diabetes Foundation.

Assistant Secretary for Health James O. Mason has accused critics of exaggerating the potential of fetal cell transplantation. Yet although the techniques are still experimental and involve small numbers of patients, the clear consensus of researchers is that the results with Parkinson's disease, at least, are encouraging. And scientists are optimistic that the findings have broader application. Fetal cells not only thrive in a host, they also can develop into specialized forms and secrete substances that nurture surrounding tissue. In addition, some fetal cells seem less likely than other cells to trigger an immune reaction.

These properties could make fetal cells uniquely valuable for treating diseases in which tissue degenerates. Parkinson's disease, for example, slowly deprives its victims of the power of movement by destroying cells in a small region of the brain. Most of the 100 or more Parkinsonian patients worldwide who have had fetal brain cells implanted are reported to show signs of improvement, according to Anders Björklund, a Swedish pioneer in the field.

Björklund readily agrees that the technique must be refined. Thirty-four months after transplantation, positron emission tomography scans indicated that only 5 to 10 percent of the cells he and his colleagues implanted into one group of four patients appeared to have survived. Even so, three of the four patients show improved mobility, Björklund says. Similarly, Robert E. Breeze, who has performed fetal tissue transplants in Parkinsonian patients with Curt R. Freed of the University of Colorado, says seven of eight patients are improved, one dramatically. Breeze and Freed support their research with private funds and patient fees.

Parkinson's is not the only disease that might be treatable with fetal tissue. Several researchers at Pittsburgh reported progress using pancreatic tissue grafts to reduce diabetics' reliance on insulin. Using fetal tissue could eliminate some of the problems encountered in obtaining tissue from adult cadavers. The one U.S. researcher who has achieved promising results treating diabetics with fetal pancreatic tissue, Kevin J. Lafferty of the University of Colorado, discontinued the work



Poll: Fetal Cell Experiments Should Continue

A majority of the U.S. public does not agree with attempts by the antiabortion lobby and the Bush administration to hold back research on fetal tissue transplantation. A poll by Louis Harris and Associates found that 74 percent of those surveyed believe experiments should continue, including 55 percent of those who oppose abortion. The survey of 1,255 adults was conducted in June. Here are the results:

	Total (percent)	Antiabortion	Proabortion*
Do you believe that research in this field should continue or not?			
Yes, should continue:	74	55	87
No, should not continue:	22	42	9
Not sure:	4	3	3
Allowing fetal cell research using aborted fetuses should be forbidden because it will encourage more abortions.			
Agree:	33	60	17
Disagree:	64	37	81
Not sure:	3	3	2
The use of aborted fetuses for research should be allowed if the alternative source—chiefly miscarriages—is inadequate in volume and quality.			
Agree:	61	33	79
Disagree:	33	61	17
Not sure:	6	6	5

*Totals may not equal 100 because of rounding.

when the ban came into effect in 1988.

Another disease that might one day be treatable with fetal tissue is muscular dystrophy, according to Terrence A. Partridge of Charing Cross and Westminster Medical School, London. And Paul R. Sanberg of the University of South Florida implants rat fetal tissue in rat brains that have chemically been damaged to cause symptoms resembling those of Huntington's disease, another devastating neurodegenerative disorder. Sanberg says his treatment can restore normal levels of activity. Deane B. Jacques of the Hospital of the Good Samaritan in Los Angeles, who plans to use private funds to treat Parkinsonian patients, hopes within a few years to experiment with treating patients disabled by cerebral palsy and stroke in the same way.

The encouraging results serve only to rankle further the handful of U.S. scientists who have proceeded despite the ban. J. William Langston of the California Parkinson's Foundation in San Jose, who has sent brain-damaged patients to Sweden where Björklund's colleague Stig Rehncrona implants fetal cells, worries that the administration's ban is eroding the quality of fetal transplantation research by forcing it to go outside the peer-review system of the National Institutes of Health. Breeze argues that the ban slows progress toward determining the best age for transplanted tissue and impedes efforts to determine where precisely such tissue should be placed.

Other countries have moved ahead by taking measures to assure that the research does not encourage abortion, either by inspiring altruism in potential donors or creating a market for fetuses. In England, where Edward R. Hitchcock of the University of Birmingham has performed several dozen fetal transplants in Parkinsonian patients, apparently with beneficial results, women who have already decided to have an abortion are asked if they are willing to donate fetal tissue for medical research. Donors are not told what disease is being studied, in order to avoid evoking sympathy on their part.

Ironically, such safeguards parallel those proposed by the U.S. review committee in 1988. It insisted that financial inducements should be illegal and that women should not be able to donate fetal tissue for named beneficiaries. In addition, the committee said women should be asked to donate only after they had made a firm decision to have an abortion. Among those voting with the 17–4 majority on the committee was Bernadine Healy, who is now director of the National Institutes of Health.

Otis Bowen, who as secretary of health under President Reagan enacted what was then seen as a temporary ban, said recently that he thought the government should have permitted federal support of the research once it received the advice. But the present administration still contends that such guarantees might not work. "It's very unlikely [donors] will not know the humanitarian effect of donating tissue," said Assistant Secretary Mason at a recent press conference in Washington. "This may well be part of their decision about whether or not to have an abortion." (Mason also admitted there are no data to support that view.)

In an attempt to defuse mounting opposition, the administration announced in May that it would spend $3 million to establish "banks" of fetal tissue from "spontaneous abortions"—miscarriages—and from ectopic pregnancies, in which a fetus develops outside the uterus and has to be removed to save the mother. Mason said transplantation research using tissue from the proposed banks would not be "encouraging or justifying elective abortions," and Healy declared the approach to be "feasible." But the researchers knowledgeable about fetal development protest that the proposed banks are impractical. "As far as I know, there is not anybody who is involved who believes this is a reasonable proposal," Redmond says.

Redmond points out that researchers have avoided tissue from miscarriages and ectopic pregnancies for good reasons: tissue for transplantation should be kept in sterile controlled conditions, and its developmental stage must be known. That is possible in a medical facility, but fetuses spontaneously aborted are likely to become infected. Indeed, a high proportion of miscarriages occurs because of infection or some genetic abnormality.

Moreover, "tissue from thousands of miscarriages would have to be collected and cryopreserved, and subsequent pathological, cytogenetic and microbiological studies would have to be done in order to obtain the 3.58 fetuses per year that might be usable" in a large hospital, Redmond maintains. "It simply would never work." Nor would using fetuses obtained from an ectopic pregnancy, which is a surgical emergency: delaying the procedure to get informed consent and to assemble a team to preserve the tissue would be unethical and dangerous, Redmond says. And tissue from ectopic pregnancies is often nonviable because it has a poor blood supply.

Healy has suggested that soon it will be possible to culture fetal cells in the laboratory. If so, cells from a single ectopic pregnancy might then be used to treat hundreds of patients. And many investigators are working on ways to implant animal cells into patients without triggering immune reactions. But those techniques are not yet available. In the meantime, according to Breeze, the administration's position is "greatly" hindering progress toward treatments that could relieve enormous suffering. —*Tim Beardsley*

DINNER IS SERVED: caterers prepare crickets, mealworms and other delicacies for an entomological feast. Photo: Jim Lukoski/Black Star.

Entomophagy

A meal of cooked insects offers food for thought

Entomologists and their guests were gathering on the patio of the Explorers Club in Manhattan to snack from a large bowl of roasted crickets and larvae. With every handful removed, the remaining contents shifted in a way that looked distressingly animated. Meanwhile Gene DeFoliart, professor emeritus from the University of Wisconsin, answered an inevitable question. "Americans say, 'I could never eat a cockroach,' as if cockroaches were the only insects," he remarked, then ticked off why nobody—anywhere—eats cockroaches: they are filthy, allergenic and chock full of bacteria.

When the New York Entomological Society celebrated its centennial this past May with a banquet, the organizers decided to put insects on the menu: chocolate cricket torte, mealworm ganoush, wax worm fritters with plum sauce and more. The society also invited DeFoliart, the editor of the *Food Insects Newsletter,* to deliver an after-dinner talk on the merits of entomophagy.

According to DeFoliart, the squeamishness that many Westerners feel about eating insects is a minority view. Outside of Europe and North America, most people throughout the world eat at least some insects. (No one knows why European cultural cuisine became so fastidious.) At movie theaters in Colombia, for example, roasted ants are eaten like popcorn. Insects are a major staple in some rural areas: one study found that in parts of Africa, more than 60 percent of the dietary protein comes from insects.

A grasshopper may look less appetizing than a drumstick, but insect bodies are nutritious. When dried, their protein content per pound is higher than that of conventional meat, although much of this protein is in the form of indigestible chitin, the material of the exoskeleton. Grubs and caterpillars often contain an abundance of unsaturated fats. Many insects are also rich in vitamins and lysine, an amino acid deficient in the diets of many people who subsist primarily on grain, DeFoliart says. Moreover, people do not just eat insects out of necessity—they enjoy the flavors.

"*Chacun à son goût,*" some cynics might shrug, but DeFoliart argues that such narrow-mindedness is harmful. Because Americans and Europeans do not consider insects a food source, few funds are available for research on the food uses of insects. Such research could help identify particularly nourishing insects and improve methods for raising and harvesting them.

As a consciousness-raising event, the banquet was successful. The only off-putting items were the three-inch, sautéed Thai water bugs, which were entirely too similar to giant cockroaches. The honeypot ants, which had transparent abdomens distended with peach nectar, were delightful sweets. And aside from their shape, the roasted crickets were almost indistinguishable from salted peanuts. As DeFoliart joked during his talk, the genus name of crickets is an open invitation to dine: *Gryllus.* —*John Rennie*

Counting Down

Pressure mounts on the U.S. to stop nuclear weapons test

On April 8 the French government took a step that would have been unimaginable before the end of the cold war. It announced it was suspending through at least the end of this year its nuclear weapons tests, which since the early 1960s have taken place in the South Pacific. President François Mitterand urged the U.S. to do likewise, arguing that such an action would help stem the spread of nuclear weapons to other countries.

On April 30 the U.S. responded to the French plea by detonating a nuclear bomb in a tunnel at the U.S. Department of Energy's Nevada Test Site. The blast—a so-called effects test, which investigates how military hardware, including nuclear warheads, will perform in a "nuclear environment"—was the second this year and the 30th since George Bush became president in 1989.

Just as Bush stood alone among heads of state attending the United Nations Conference on Environment and Development in refusing to sign the biodiversity agreement, so has he become increasingly isolated on the issue of nuclear testing. Now that the cold war is over, proponents of a test ban say, the U.S. has nothing to lose and much to gain from ending its tests. "The discontinuance of nuclear testing is seen worldwide as a first step toward a world of zero nuclear weapons," says Frank von Hippel, a physicist at Princeton University and an authority on arms control. "If we insist that nuclear weapons are forever, then, over time, more and more countries will argue that they too need such weapons."

Arms-control advocates hoped Bush might take a small step toward a test ban in June during his summit meeting with Russian president Boris N. Yeltsin. Military leaders in Russia had been pressing Yeltsin to end that country's unilateral moratorium on testing, and to keep the hard-liners at bay Yeltsin had reportedly sought some sign that the U.S. was committed to a ban—an agreement to limit the number of tests per year, for example. Although the leaders did agree to slash their missile arsenals, Bush conceded nothing on nuclear testing.

Some congressional leaders, notably Sam Nunn of Georgia, chairman of the Senate Armed Services Committee, view Bush's position as a betrayal of a pledge made by his predecessor, Ronald Reagan. In 1986 Nunn and others agreed to squelch a test-ban bill and to sup-

port Reagan's efforts to complete negotiations with the Soviet Union on the Threshold Test Ban Treaty, which prohibits tests with yields exceeding 150 kilotons of TNT. In exchange, Reagan said the U.S. would "immediately" begin negotiations toward a comprehensive test ban after the threshold treaty was completed.

The Soviet Union and the U.S. had been voluntarily abiding by the 150-kiloton limit (roughly 10 times the yield of the bomb that destroyed Hiroshima) since 1974, but their inability to agree on verification measures prevented them from making the agreement official. In 1990 the Soviet Union finally accepted all the U.S. demands on verification, which included U.S.-designed sensors and on-site inspections. Both nations signed the treaty, and the U.S. Senate ratified it.

Congressional leaders then waited in vain for Bush to fulfill the promise made by Reagan. This past March, Nunn and two other Democratic senators, Al Gore of Tennessee and Paul Simon of Illinois, wrote a letter to the president's national security adviser, Brent Scowcroft, reminding him of Reagan's "solemn commitment." "We cannot overemphasize," the letter stated, "the implications for possible legislative action by Congress in the nuclear testing area this year were the Bush Administration to renege on this commitment."

Indeed, Congress is now seeking to take matters into its own hands. In early June the House of Representatives voted 237–167 for a bill imposing a one-year moratorium on testing. A similar bill, still pending, has been sponsored by almost half of the members of the Senate. But even if the bill is passed by both houses of Congress, the president is almost certain to veto it.

Obviously, the U.S. can no longer maintain that it must develop and test new warheads to keep pace with the now defunct Soviet Union. And in fact, tests of advanced warhead designs, such as the exotic nuclear-pumped x-ray laser warhead once being considered for the Strategic Defense Initiative, have been drastically cut back. But Secretary of Defense Dick Cheney and other officials insist that tests are still needed to ensure that weapons already in the stockpile are reliable and safe—in other words, that they will explode when called on to do so but not sooner.

Those justifications have been rebutted by Ray E. Kidder, a physicist at Lawrence Livermore National Laboratory. Kidder, who has 35 years of experience as a designer of nuclear weapons, says nonnuclear tests are more than adequate to ensure that warheads will blow up only on command. As for safety, he notes that most U.S. weapons already incorporate advanced, fully tested safety features, such as low-sensitivity explosives to initiate the blast.

Of those weapons that lack such features, only three are not scheduled for immediate retirement: the Minuteman 3 ground-based missile and the Trident 1 and 2 submarine-launched missiles. The U.S. can modernize these missiles simply by replacing their warheads with more advanced models, according to Kidder. No more than 12 underground tests, he says, would be required to demonstrate the reliability of the modified weapons.

Even these few tests may be unnecessary, says Matthew Bunn of the Arms Control Association. He points out that both the U.S. Air Force, keeper of the Minuteman 3, and the Navy, which oversees the Trident missiles, have stated that the weapons are already safe and do not need modifying. The advantages of having an immediate test ban, Bunn says, greatly outweigh the benefits of further safety-related tests.

As if to underscore this point, China on May 21 conducted its most powerful nuclear test yet. Although the Chinese blast, which had an estimated yield of more than 1,000 kilotons of TNT, was conducted underground, it reportedly released a large plume of radioactive gas.

Now that France has suspended its tests, Bunn points out, China is the only nation besides the U.S. and Great Britain that is openly conducting tests and is not publicly committed to a comprehensive test ban. (Great Britain conducts roughly one test every 18 months at the Nevada Test Site and would therefore probably cease testing if the U.S. did.) Bunn argues that by halting U.S. tests and applying some "strong diplomacy," Bush could almost certainly persuade China—the country to which he was once ambassador—to join a test ban. "China would find it embarrassing to oppose a comprehensive test ban at this point," Bunn says.

Kidder, the Livermore weapons designer, predicts that a test ban will be achieved within four years, even if Bush is reelected. Although "Bush and a few diehards are still hanging on," Kidder says, even the weapons laboratories are losing their enthusiasm for testing now that the emphasis on designing new warheads has declined. "When I was a lad here" in the late 1950s, Kidder recalls, "we were banging them off in the atmosphere, out in the South Pacific. It was very glamorous, and we were designing new things. Now it's a dead issue." —*John Horgan*

Plenty of Nothing

A detector finds more particles than expected—but not enough

When physicists began operating the Gallium Experiment in the Gran Sasso tunnel below the Italian Apennines a year or so ago, they had high hopes. The primary purpose of GALLEX, which consists of a tank containing 50,000 liters of a solution laced with the element gallium, was to observe the elusive, possibly massless particles known as neutrinos and thereby resolve the so-called solar-neutrino problem.

The problem is this: the standard model of nuclear physics yields a remarkably precise estimate of the number of neutrinos spawned by nuclear fusion in the sun's interior, but observations made during the past 24 years by three different neutrino detectors have come up far short. The discrepancy led some workers to call for revisions in the standard model of particle physics, while others argued that the data were tentative and needed confirmation [see "The Solar-Neutrino Problem," by John H. Bahcall; SCIENTIFIC AMERICAN, May 1990].

Now that the first year's results from GALLEX have been reported, they muddy rather than clarify the situation, according to Richard Hahn of Brookhaven National Laboratory, a member of the GALLEX team. If the detector had observed more neutrinos, he explains, "everybody would say, 'Aha, they're confirming the standard model.'" Results that were lower, nearer to those from the previous experiments, would have confirmed the need for a new physics model. "But we're sort of in-between," Hahn says.

Like previous experiments, GALLEX detects only an infinitesimal fraction of the vast numbers of solar neutrinos that pass through it every second. Every now and then, a neutrino strikes a gallium nucleus and transforms it into a radioactive isotope of germanium. By measuring the radioactive germanium, the GALLEX workers have estimated that roughly two thirds as many neutrinos are passing through the detector as predicted by conventional models of nuclear and solar physics.

That percentage is still significantly higher than those derived from earlier experiments. Workers began monitoring solar neutrinos in 1968 with a vat of a chlorine compound called perchloroethylene—ordinary dry-cleaning fluid—in a gold mine in South Dakota. (Neutrino observatories are placed un-

derground to shield them from cosmic rays that could produce spurious signals.) When a neutrino collides with a chlorine nucleus, it is transmuted into a radioactive isotope of argon. After 24 years of measuring argon isotopes, workers found less than a third of the predicted number of neutrinos.

In 1986 a more direct type of observatory consisting of a large tank of water began operating in a mine near Kamioka, Japan. The collision of a neutrino with an electron in the water releases a flash of light, called Cerenkov radiation, that is detected by photomultiplier tubes surrounding the tank. So far the experiment, which is called Kamiokande II, has detected about 40 percent as many neutrinos as predicted by the standard model.

The lowest numbers of all have come from a detector in Russia's North Caucasus region that, like GALLEX, consists of gallium in solution. Called (now anachronistically) the Soviet-American Gallium Experiment, or SAGE, it began operating four years ago. Unlike either the South Dakota or Kamiokande II observatories, which can detect only highly energetic neutrinos produced by relatively rare types of fusion, SAGE was designed to detect the low-energy neutrinos generated by the fusion of two hydrogen nuclei, or protons, which is thought to be the source of almost all solar neutrinos.

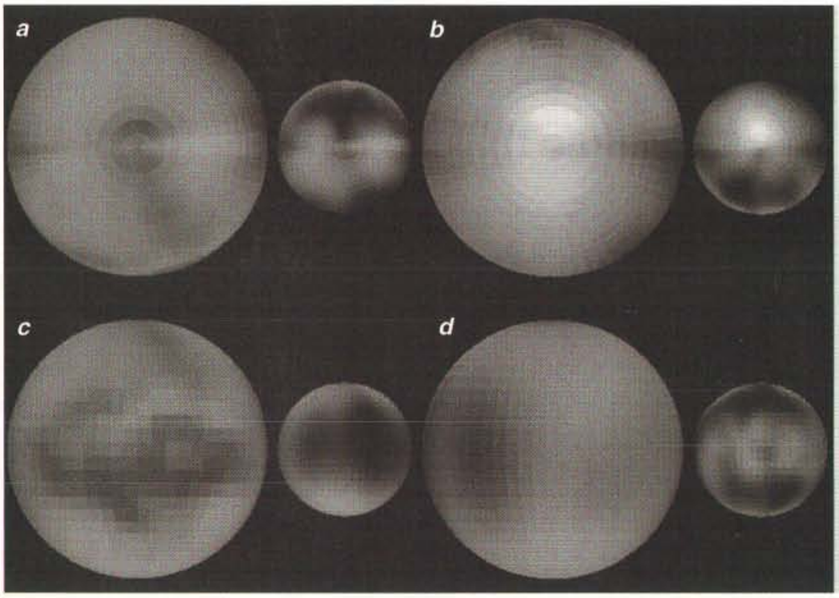
A Rare Glimpse of a Dim World

Ever since it was discovered in 1930, Pluto has stood out as a planetary oddball. Its highly elliptical orbit periodically carries it more than half again as far from the sun as any other planet. It is tiny—a mere 2,300 kilometers across, or about two thirds the diameter of the earth's moon. Knowledge about Pluto's appearance had remained scanty because the planet appears as a featureless blob even through the finest telescopes.

From 1985 to 1990, however, nature afforded astronomers a way to glimpse details on the planet's surface. During those years, Pluto and its large moon Charon, which measures 1,200 kilometers in diameter, were aligned as seen from the earth. Charon therefore appeared to pass in front of and behind the planet once every 6.4 days, the time Charon takes to complete one orbit. As Charon obscured part of Pluto's surface and as Pluto eclipsed its satellite, the combined brightness of the two objects varied slightly. Such mutual eclipses occur only once every 120 years. Eager to make the most of the opportunity, several research teams closely monitored those fluctuations in brightness and analyzed the information to locate the bright and dark patches on Pluto and on its satellite.

Two groups—one led by Marc W. Buie of Lowell Observatory and the other by Richard P. Binzel of the Massachusetts Institute of Technology—have just released the product of their years of labor: the first detailed views of Pluto and Charon. Although the teams worked independently, their maps of surface reflectivity closely resemble each other, a reassuring sign. Both maps show that Pluto is dark at mid-latitudes and distinctly brighter around the north pole. Most intriguingly, the planet has an extremely bright south polar region, which hints at the seasonal changes that sweep across this remote world.

The area around the south pole reflects 98 percent of the light that strikes, making it as white as freshly fallen snow. Binzel thinks that bright zone consists of a thin layer of methane ice. Pluto's atmosphere contains methane that probably evaporated from the surface as the planet approached its closest point to the sun, which occurred in 1989. As the planet circles away from the sun and temperatures drop, much of the methane will freeze back onto the surface. Pluto's orientation now is such that the south pole is moving into darkness and will remain there for one half of a Plutonian year, or 124 terrestrial years. Consequently, the frost should preferentially settle on the frigid, shadowed south pole.



RECONSTRUCTED IMAGES of Pluto and Charon depict the bodies as seen from the north (a) *and from the south* (b). *Marc W. Buie of Lowell Observatory and his co-workers inferred one set of hemispheres from eclipse data* (c) *and the other* (d) *from cruder observations of rotational light curves.*

Buie and his collaborators have mapped Charon in addition to Pluto; they report that Charon is darker overall and that some regions seem to reflect only about 3 percent of the incoming light, about as much as a lump of coal would. One likely reason for the dissimilarity is that Charon's gravity may be too weak to hold on to its methane and so does not develop a seasonal covering of frost.

These findings come at a time of increasing interest in Pluto and Charon. S. Alan Stern of the University of Colorado argues that Pluto and Triton, Neptune's largest moon, are the sole known survivors of a swarm of icy bodies that circled the outer reaches of the solar system during its infancy. If so, they harbor a great deal of information about how the outer planets formed and why they are so different from the earth in size and composition. Even as it recedes ever farther from the sun, Pluto is at last getting its moment in the astronomical limelight. —*Corey S. Powell*

Last December, however, the SAGE team reported that it had observed only 20 percent, at most, of the predicted number of low-energy neutrinos in its first year. Some researchers suggested that even these few apparent detections were spurious and that the experiment had actually detected no neutrinos.

The disagreement between GALLEX and SAGE is particularly vexing, since the experiments are similarly designed and are sensitive to neutrinos in the same energy range. On the other hand, SAGE workers recently reported that since they doubled the amount of gallium in their experiment last year, they have been detecting neutrinos in greater numbers. The most recent data are reportedly in rough agreement with the GALLEX findings.

John Wilkerson of Los Alamos National Laboratory emphasizes that the results from both GALLEX and SAGE should be considered preliminary, since neither has been properly calibrated yet. (Calibration involves exposing the detectors to a radioactive source whose output of neutrinos is known.) Wilkerson points out, moreover, that because the detectors built thus far record so few "events"—SAGE, for example, observed only three putative neutrinos during all of 1990—their findings are statistically suspect.

More robust data should be forthcoming from a highly sensitive neutrino detector located in a mine near Sudbury, Ontario, and scheduled for start-up in 1995. Called the Sudbury Neutrino Observatory, or SNO, it is similar to the Kamiokande II experiment but uses heavy water (formed when oxygen bonds to deuterium, an isotope of hydrogen containing a neutron as well as a proton). Sensitive to a broad spectrum of neutrino interactions, SNO should be able to detect as many as 3,000 neutrinos a year, more than 20 times the number achievable by any other experiment. "One of the keys to solving the solar-neutrino problem," Wilkerson says, "will be an experiment like SNO with very high statistics."

The payoff for solving the problem could be high. Already some physicists who think the apparent deficit in solar neutrinos is real have sought to explain it by proposing modifications to the standard model of nuclear physics. The most popular modification, known as the Mikheyev-Smirnov-Wolfenstein theory, requires that at least one type of neutrino have mass, perhaps enough to account for the mysterious "dark matter" that many astrophysicists believe dominates the structure of the universe. *—John Horgan*

Particle Goddess

A new accelerator peers inside protons

To look deep into the structure of matter, physicists will gather this month under the streets of Hamburg, Germany. There lies the Hadron-Electron Ring Accelerator (HERA), which recently began smashing electrons into hadrons, a class of heavy fundamental particles that includes protons. Like an immense microscope, HERA promises to reveal structures smaller than a billionth of a billionth of a meter, a distance 1,000 times smaller than the diameter of a proton. "The goal of HERA is to look at the proton at a very high resolution, a factor of 100 greater than people have done before," says Bjorn Wiik, who, with Gustav Voss, directed the construction of HERA.

Housed in a circular tunnel 6.3 kilometers long, HERA is designed to accelerate protons to energies of 820 GeV (billion electron volts) and electrons to 30 GeV. The machine is the most powerful collider of protons and electrons. In terms of energy alone, however, HERA has two rivals. The LEP collider at the European facility CERN accelerates electrons and positrons to energies of 50 GeV, and the Tevatron at Fermilab in Illinois propels protons and antiprotons to 900 GeV. But like the Greek goddess, HERA has special abilities of her own.

HERA can propel electrons deep into the interior of the proton. There an electron interacts with the proton's constituent particles: three quarks and a number of gluons. The three quarks contribute mass and electric charge to the proton. The gluons hold the quarks

East of Eden

These skulls from China's Hubei Province tell of a time before *Homo sapiens* had emerged. Yet their finders say the specimens tell much about that emergence. The workers cite faunal and other evidence to date the skulls to at least 350,000 years ago, long before the time when other scholars place the first modern-looking people in Africa and in Africa alone. Moreover, they say, the skulls' flat faces and high cheekbones testify to a racial continuity with modern Asians that contradicts the out-of-Africa theory.

"I wouldn't say Asia is the homeland of modern humans," says Dennis A. Etler, a doctoral candidate at the University of California at Berkeley, who reported on the skulls in *Nature* with Li Tianyuan of the Hubei Institute of Archaeology. "That goes totally against what we're trying to say, which is that modern humans don't have one exclusive place of origin."

Li found the skulls, deformed and encrusted in mineral deposits, under newly terraced ground that had already churned up other bits of fossil bone. The crushing obscured some features, and the mineral matrix still clogging the braincases prevents direct measurements of their capacity. Yet Etler says external evidence shows a brain size at the "very upper end of *H. erectus,* if not beyond it." Continuity with that earlier species is apparent, he adds, in the cranial vault, which is long, low and angled toward the back like a football.

But Christopher B. Stringer of the Natural History Museum in London continues to support an African homeland for humanity. He says the Chinese skulls may represent a dead-end branch stemming from an earlier population that some scholars call *H. heidelbergensis*. "*Heidelbergensis* could have split," Stringer says, "in Europe becoming Neanderthals, in Africa modern humans and in China—we don't know. All we can say is that well-dated modern humans do not turn up in China until 35,000 years ago. In Africa we have a chain showing the transition, and my view is that the evidence still supports the African area." *—Philip E. Ross*

together by conveying the strong force. When the electron collides with a quark, it cannot kick the quark free. But the collision will produce jets of particles that, when analyzed, will reveal the internal structure of the proton.

The accelerator will truly deserve a place on Mount Olympus if it produces an exotic particle known as a leptoquark. All matter consists of either quarks, such as those in the proton, or leptons, such as electrons and neutrinos. Physicists predict that if a quark and a lepton collide at literally lightning speeds they will bind together to form a leptoquark. If discovered, physicists will begin to understand just why nature created both quarks and leptons.

HERA took eight years to build and cost about $650 million (DM 1,010 million). It requires two independent accelerator systems, one for the protons, the other for electrons. The proton accelerator must keep the relatively heavy protons on track and therefore requires strong superconducting magnets. The electron accelerator's most important challenge consists of replenishing the energy that the particles lose in the form of synchrotron radiation. A series of sophisticated radio-frequency cavities makes up the deficit. No less complex is the feat of guiding the beams of electrons and protons so that they collide in the center of two house-size detectors.

In May, HERA's two large detectors recorded their first events, and in the fall, researchers expect HERA will begin to show her true power. *—Russell Ruthen*

Sea Sick

Marine mammals are a barometer of ocean health

The big-eyed seals and sea lions lolling at the Marine Mammal Center are drawing murmurs from visitors—some because the animals are cute, others because they are so obviously ill. But determining what exactly is wrong and the best kind of care to administer is difficult. "When humans walk into an emergency room, we have a very solid idea of what their immunologic profiles should be, but we have only limited baseline data for marine mammals," points out Peigin Barrett, executive director of the private hospital and rehabilitation center in Sausalito, Calif.

Every year the center assists hundreds of mammals that wash up along 1,000 miles of California coastline. These stranded creatures are "barometers of what's going on in the oceans," Barrett declares. "We should be studying them not just to make a few animals feel better but to get a balanced look at the ocean environment we humans depend on and share with other countries." The unique research opportunity that stranded mammals represent has been largely missed, she observes: "The technology has been there. The dollars have not."

Yet Barrett's center and others could soon receive a financial boost if legislation set to go before Congress this summer is approved. A new bill, an amendment to the Marine Mammal Protection Act of 1972, would corral funds to study the health of these creatures and to coordinate effective responses to strandings and catastrophes.

Impetus for supporting systematic study of marine mammals is growing as die-offs and massive strandings become more frequent, asserts John R. Twiss, Jr., executive director of the Marine Mammal Commission in Washington, D.C. The most recent spate of deaths began this spring, with the appearance of distemper virus in seals along the northeastern coast of the U.S. A die-off of bottlenose dolphins that started this past winter continues off the coast of Texas. Over the past two years, a virus has killed hundreds of striped dolphins in the Mediterranean. "There are clearly things going on with marine mammal health that are not well understood," Twiss says.

Several factors are probably involved in the massive mortalities, suggests Joseph Geraci, an expert on marine mammal deaths at the Ontario Veterinary College in Guelph. Stress from high populations and competition with commercial fisheries may enhance mammals' susceptibility to infectious viruses, Geraci says. "That's not to discount the possible effects of numerous contaminants animals are eating every day and accumulating in blubber," he adds.

Geraci says he has not yet encountered evidence to convince him that man-made pollutants are an underlying cause of the die-offs. Biological poisons produced by microscopic aquatic organisms have been involved in at least two major outbreaks and perhaps more, he notes.

It is impossible to forge any causal links without a reliable data base, says Nancy Foster, director of the office of protected resources at the National Marine Fisheries Service. She pushed for the founding of the National Marine Mammal Tissue Bank, established in 1988. The amount of data is growing as stranding networks around the country follow systematic protocols for collecting and preserving tissue. "We want people to be able to rely on this in 20 or 100 years," Foster says.

Whales and dolphins may well be sentinels for ocean health, but they are also totems for public admiration. The double-edged sword of public sentiment sometimes slices at researchers attempting to take blood samples from beached animals. The cry goes up, "Let them die in peace," even as workers try to explain that the animals' deaths will not be in vain if they provide understanding. *—Deborah Erickson*

PEIGIN BARRETT of the Marine Mammal Center, a private rehabilitation center, aids hundreds of stranded and dying sea creatures every year. Photo: Stephanie Rausser.

PROFILE: PHILIPPA MARRACK AND JOHN KAPPLER

Keeping It in the Family

In a city park two blocks north of their Denver laboratory, while their co-workers finish the last of a picnic lunch, Philippa Marrack and John Kappler pace off 100 yards for a footrace. He walks with confident, measured steps, eyes to the ground, alert to small inaccuracies. She has assumed a more lighthearted gait, her arms and legs swinging jauntily in what might be a parody of a British lieutenant's march. "We ran the 100 yards the other day, and John was the fastest person in the lab," Marrack explains. Kappler grins. "I was the fastest woman," she continues, "but then, I was the only one running."

In step and side by side, Marrack and Kappler have been winning honors in immunology for two decades. As a team, they have done much to explain how the immune system balances on the razor's edge between self-tolerance and autoimmunity—acceptance and rejection of the body's own tissues. They are both investigators for the Howard Hughes Medical Institute at Denver and members of the elite National Academy of Sciences. They are also, perhaps not so incidentally, husband and wife.

"The scientific world always considers us a single word: Kappler-and-Marrack," she says over a bowl of soup in the cafeteria of the National Jewish Center for Immunology and Respiratory Disease. From the other side of the table, Kappler nods his agreement. He had made almost the identical point earlier upstairs in the laboratory. "Our lives are so tangled up in the lab that we don't leave it. It's part of our life at home," he said. "It's certainly an advantage that's played a big part in our success."

Nothing about their partnership or their success would have been easy to predict. As a student at a girls' high school in England during the early 1960s, Marrack reflects, "I don't think it ever occurred to me that I would be a career scientist." She nonetheless indulged her aptitude for science all the way through undergraduate and doctoral degrees in biology at the University of Cambridge. "I was just following what I was good at, I guess," she observes. "If I'd met with failure at any of those steps, I'd have stopped." Marrack says she made her decision to do postdoctoral work at the University of California at San Diego in 1971, after she and her friends decided in a pub that San Diego was the place in the world least like Cambridge.

PHILIPPA MARRACK AND JOHN KAPPLER are partners in both immunology and matrimony. Photo: Geoffrey Wheeler.

Kappler, who was born in Baltimore in 1943, was a little more gung ho in his early years. Going to high school at Baltimore Polytechnic Institute during the Sputnik era, he was encouraged to pursue his interests in science and math. Kappler headed to Lehigh University intending to be an engineer but says he was "scared off" by the first-year courses and switched to chemistry instead. After graduating from Lehigh in 1965, he followed his adviser's suggestion to go to Brandeis University, and there he became involved in the biochemistry of cell differentiation.

In 1971, after finishing his doctorate at Brandeis, he joined the laboratory of Richard W. Dutton at U.C.S.D. Dutton had recently learned how to grow cultures of normal spleen cells, including the *T* lymphocytes that control the immune system. "This was just when immunology was getting ready to explode," Kappler recalls.

Marrack and Kappler worked together in Dutton's laboratory for two years. Then, in 1973, Kappler joined the junior faculty of the University of Rochester and asked Marrack to go with him. They married in 1974. "When we went to Rochester, I had a job and Pippa didn't," Kappler explains. "We just showed up, more or less. Pippa went to work in the lab, and we organized a postdoctoral salary for her."

Establishing Marrack as a coequal member of the partnership rapidly became a priority. Fortunately, they received help from the chairman of the department, Frank E. Young, who later became head of the Food and Drug Administration in the Reagan and Bush administrations. "He recognized something good, and Pippa got an appointment pretty quickly," Kappler says.

To build her credibility further, Marrack immediately sought and won from the American Heart Association an investigatorship to do basic research. "The university could see that somebody out there thought I wasn't just John's technician," she chuckles. Kappler's attitude also encouraged her. "Very often, you'll see husband-and-wife pairs in which the husband automatically becomes the one who goes to all the meetings and gives all the talks," Marrack says. "John never let that happen."

While in Rochester, Kappler and Marrack began to study how *T* cells recognize antigens, the substances that immunologically distinguish one cell from another. If a highly specific receptor molecule on a *T* lymphocyte binds with

an antigen, the lymphocyte identifies the antigen-bearing cell as an alien intruder and attacks it. Marrack and Kappler's work helped to establish a conceptual breakthrough in immunology: *T* cells can recognize only antigens displayed by certain histocompatibility surface proteins.

Marrack and Kappler continued to make important discoveries in Rochester and later in Denver, where they relocated in 1979 to the National Jewish Center and the University of Colorado Health Sciences Center. Using antibodies, they described features of the antigen receptors on *T* cells that are critical to normal function. More recently, the two of them identified the class of substances they named superantigens.

Superantigens stimulate *T* cells far less selectively than ordinary antigens do. To turn on a *T* cell, an antigen must precisely match five different "keys" on the receptor molecule. Superantigens accomplish the same feat by matching just one key. In 1990 Marrack and Kappler showed that many bacterial molecules, including the toxins that cause staphylococcal food poisoning and toxic shock syndrome, are superantigens. Marrack went on to prove in 1991 that superantigens found in mice were actually viral in origin.

The jewel in Kappler and Marrack's crown is their 1988 proof that a mechanism called clonal deletion is largely responsible for making the immune system tolerant of the self. Using antibody probes and genetically engineered mice, they demonstrated that the immune system subjects itself to a rigorous test. Within the thymus gland, immature *T* cells are presented with a spectrum of self-antigens from throughout the body. Any clones of *T* cells that show a predilection for attacking the self are quickly culled from the ranks before they can do any harm. "That's the most important discovery we ever made," Marrack nods.

Such discoveries might not have been possible if Kappler and Marrack had not so thoroughly integrated their research with their lives. From his office, Kappler points six blocks away to the tree-lined neighborhood where he and Marrack live. Her parents also live nearby: they came to the U.S. in 1978 to help care for Marrack and Kappler's two young children. Today their daughter, Kate, is a junior in high school, and their son, Jim, is heading for college at—surprise—U.C.S.D.

All this togetherness might be expected to strain any partnership, but both scientists assert they have very natural rules that they follow to minimize professional conflicts. For example, on research publications, the person who performed the principal experiments is always the first listed author; the one who primarily wrote the paper is the last author.

"Professionally, our egos and goals and points of view clash quite often," Kappler explains, "but as far as the operation of the lab and the sharing of credit goes, we worked out how to do that long ago." Of course, as Marrack jokes, "You have to understand that one of us is a lot more secure and doesn't need to vent his ego. So that helps."

Science treats them as a single word: Kappler-and-Marrack.

Their complementary skills also lend strength to their collaboration. "I come from a chemistry background, which sometimes gives me an advantage," Kappler says. "I think Pippa does a better job on the biological sorts of experiments. I don't think we've ever worked side by side on the bench."

The inherent risk of partnerships built on complementary strengths is that the partners may decide to go in dissimilar directions. To some extent, that is happening to Kappler and Marrack. They continue to participate fully in each other's research, and both of them are still interested in the same fundamental question: How do *T* cells recognize antigens? Nevertheless, they are beginning to look at that problem from more distinct vantage points. "Now, more than ever, we have different interests," Marrack notes.

Kappler is increasingly interested in how the structure of the *T* cell's antigen receptor affects its function. In recent years, he has studied the receptor by examining how mutations affect it. Yet he knows that a real picture of the receptor will come out of another specialized field, protein crystallography. "I don't want to pull up short of the answer," he says. "With the skills I have right now, I can't take that last step and actually see the structure." For that reason, he is now trying to learn some basic crystallography.

For her part, Marrack has more cellular interests. Human autoimmune diseases, such as rheumatoid arthritis and some types of diabetes, fascinate her these days. The gulf between the current understanding of *T* cells and the pathological manifestations of disease is wide, she says. The unknowns of autoimmune disease are close at hand for Marrack: her own brother has ankylosing spondylitis, an autoimmune disease affecting the joints. She would have the condition as well, she says, except that for mysterious reasons, it develops less often in women than in men.

As she and Kappler explain, immunologists once thought that autoimmune disease followed inevitably from a failure of tolerance, such as self-reactive clones of *T* cells slipping past the screening mechanisms. More recent research has shown that, in fact, everyone seems to have some potentially self-reactive *T* cells but that they are generally dormant. The riddle of autoimmunity therefore often hangs on what activates those cells.

"I have become increasingly convinced that all the autoimmune diseases do not have the same mechanisms, that tolerance is breaking down in different ways," Marrack says. "That's why I think mouse models for human diseases may not necessarily be as useful as was once thought." Elaborating on her point, her partner says, "The mice models may be good for each individual mechanism, but diabetes in mice might be a good model for a completely different autoimmune disease in humans. It's not necessarily a one-for-one match."

An autoimmune disease about which Marrack and Kappler have been making intriguing discoveries is rheumatoid arthritis. When the couple looked at the fluid inside patients' arthritic joints, they found an overabundance of certain *T* cell clones, some of which may have been self-reactive against cartilage or some other substance in the joints. Much more to the researchers' surprise, they also observed that a large subpopulation of *T* cells was completely missing from the blood.

The meaning of these results is still uncertain. One possibility, Marrack notes, is that these patients harbor a superantigen. Although superantigens seem to activate *T* cells in the short run, they may eventually kill or neutralize the cells. A superantigen might have turned on dormant, self-reactive *T* cells in the patients; some of the reactive cells might have survived the debilitating effects of the superantigen and continued their attack on the joints.

The fusion of their professional and private lives will certainly help Marrack and Kappler in their attempt to understand autoimmunity. "There's another advantage," Marrack realizes suddenly. "If John says he wants to spend the entire weekend working at the lab, that would be a problem for some couples because the wife might say, 'You should be home with the family doing such and such.' But I reckon he's in the lab working for me." —*John Rennie*

SCIENTIFIC AMERICAN

Dynamics of Kilauea Volcano

Kilauea is one of the most thoroughly studied volcanoes in the world. That scrutiny is helping scientists to understand how volcanoes work and to predict where other destructive eruptions might occur

by John J. Dvorak, Carl Johnson and Robert I. Tilling

Shortly after midnight on January 2, 1983, we and our colleagues at the Hawaiian Volcano Observatory witnessed the beginning of one of the longest and most extensive volcanic eruptions in recorded history. A flurry of earthquakes and a steady swelling of the ground around Kilauea volcano signaled that the earth was slowly ripping apart as molten rock incrementally forced its way toward the surface. Almost exactly 24 hours after the initial burst of activity began, a red glow appeared on the eastern horizon. At the same time, the earthquakes ceased and were replaced by a rhythmic vibration—the telltale heartbeat of lava gushing from the ground.

Kilauea continues to erupt to this day. Since 1983, its lava flows have covered nearly 100 square kilometers, roughly the area of Manhattan, and added 120 hectares of new land to the island of Hawaii. Kilauea's latest outpourings have destroyed more than 180 homes and displaced hundreds of people.

The outbursts have also had important salutary effects. Study of Kilauea's behavior has helped scientists to predict eruptions and thereby to ameliorate the loss of life and property caused by other volcanic events, such as the eruption of Mount Pinatubo in the Philippines in 1991. Moreover, the volcano offers a unique window into the workings of the earth's interior. Geologists' deepest drill holes penetrate little more than 10 kilometers into the earth,

JOHN J. DVORAK, CARL JOHNSON and ROBERT I. TILLING have all worked at the Hawaiian Volcano Observatory. Dvorak is now a researcher at the Cascades Volcano Observatory in Vancouver, Wash. He specializes in measuring ground motions around active volcanoes. Johnson holds the position of professor of geology at the University of Hawaii at Hilo; during the summer, he teaches at the university's Center for the Study of Active Volcanoes, an institution that trains scientists from other countries on techniques of monitoring volcanoes. Tilling is a researcher at the U.S. Geological Survey's branch of igneous and geothermal processes in Menlo Park, Calif. He oversaw the scientific response to the explosive eruption of Mount St. Helens in 1980 and has assisted many international programs for improving volcano monitoring and reducing volcanic risk. Tilling also directed the Hawaiian Volcano Observatory during the 1970s, when Mauna Loa reawakened after a 25-year slumber.

LAVA FLOWS from Kilauea's current eruption have transformed the local environment on the island of Hawaii. From 1983 to 1986 the volcano spouted spectacular episodic fountains of lava, as seen in the distance in this 1985 photograph. Molten rock now issues more calmly and steadily from the ground. Such activity is the latest manifestation of an enduring geophysical pulse that built the entire chain of Hawaiian volcanoes over the past 70 million years.

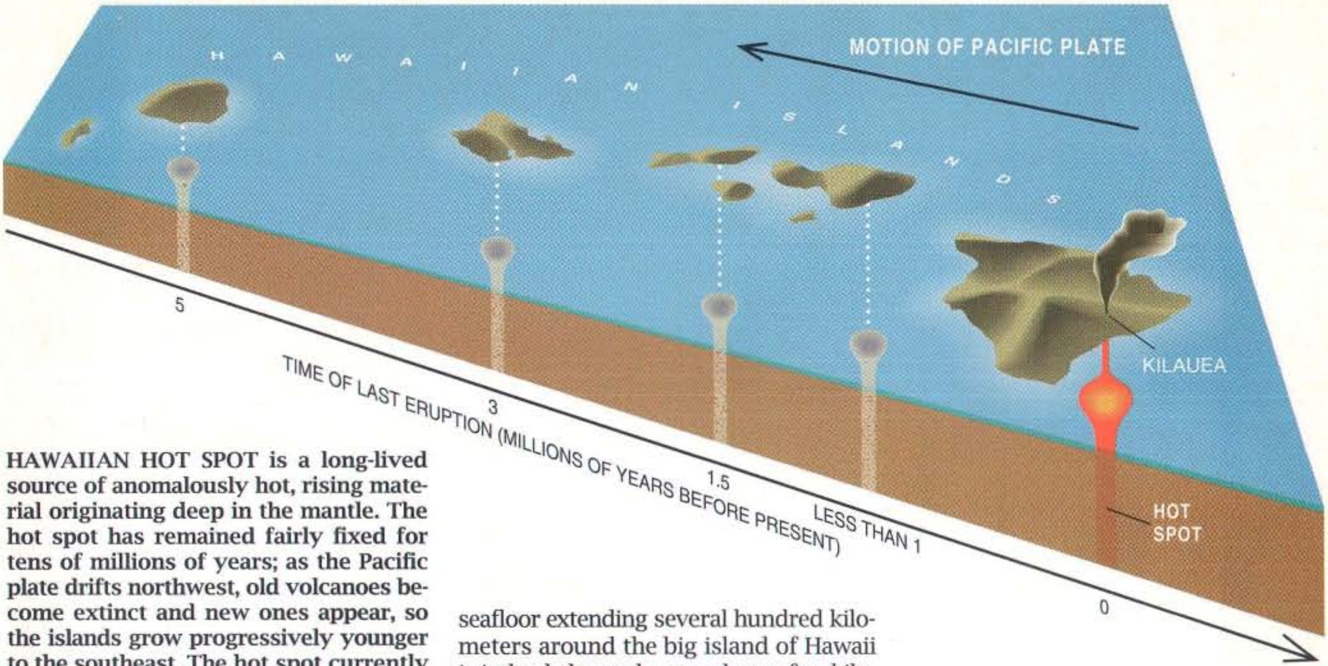
only one six-hundredth the way to the center. Lava that has emanated from Kilauea provides a direct sample of material from tens or perhaps even hundreds of kilometers inside the earth. Each eruptive episode presents an invaluable showcase of the volcanic processes that have formed more than 80 percent of the earth's total surface above and below the sea.

Kilauea is the most vigorous of the five volcanoes that make up the large island of Hawaii. Kilauea's eruptions have been watched for many centuries by the Polynesian inhabitants of the chain. They gave the volcano its name, which means "rising smoke cloud." The native Hawaiians attributed volcanism to the actions of the goddess Pele. Her traditional home is Halemaumau ("house of the everlasting fire"), the large pit crater at the summit of Kilauea. Halemaumau has been a site of recurring volcanic activity at least since the first written descriptions of Kilauea, which date to the 1820s.

Hawaiian mythology suggests that the early Polynesian inhabitants possessed a surprisingly profound understanding of the islands on which they lived. According to legend, the Hawaiian islands grow progressively older from southeast to northwest. Modern dating techniques, which require precise measurement of radioactive elements in rock samples, confirm that age progression. The big island of Hawaii is the youngest island in the chain. It has formed within the past one million years and is still growing. The most ancient volcanoes of the chain, located some 5,000 kilometers to the northwest, near the Aleutian islands, are approximately 70 million years old.

The orderly, progressive ages of the Hawaiian islands and their location in the middle of the Pacific Ocean at first seem rather puzzling. Most of the earth's geologic activity is thought to be associated with the motions of large blocks of crustal and upper mantle rocks, many of which outline the major continents. These blocks, or plates, lie over hotter, more pliable rocks deeper in the mantle. Volcanoes and powerful earthquakes tend to occur near the boundaries between plates, where the earth's surface compresses or pulls apart. The famous "ring of fire" encircling the Pacific Ocean, which encompasses the volcanoes in the Philippines, Japan and Alaska, marks the edge of the huge Pacific plate. But Kilauea sits in the middle of this plate, far from any such locus of activity.

The mystery of Kilauea and other, analogous midplate volcanoes was explained in 1963 by J. Tuzo Wilson of the University of Toronto [see "Continental Drift," by J. Tuzo Wilson; SCIENTIFIC AMERICAN, April 1963]. Wilson theorized that the age pattern of the Hawaiian islands attests to the slow march of the seafloor over a deep-seated, relatively fixed upwelling of molten rock in the earth's mantle, which he termed a "hot spot." As the Pacific plate drifts to the northwest, he postulated, the hot spot creates a linear succession of volcanoes. Geophysicists have now



HAWAIIAN HOT SPOT is a long-lived source of anomalously hot, rising material originating deep in the mantle. The hot spot has remained fairly fixed for tens of millions of years; as the Pacific plate drifts northwest, old volcanoes become extinct and new ones appear, so the islands grow progressively younger to the southeast. The hot spot currently feeds magma to five volcanoes on the island of Hawaii, including Kilauea.

identified at least 100 possible such features around the globe, but the Hawaiian hot spot is the most vigorous and best understood.

Wilson calculated that based on the relative ages of the Hawaiian islands, the Pacific plate moves nine centimeters a year on average. Recent measurements of the relative positions of landmasses based on the timing of radio signals from deep space have confirmed that the floor of the Pacific Ocean is advancing northwest at almost exactly that rate. After roughly a million years, each Hawaiian volcano becomes extinct as the Pacific plate ferries it away from its magma source. The magma from the hot spot then surfaces at an adjacent location, initiating the formation of a new volcano. The entire chain of islands was presumably created in this manner.

In 1971 W. Jason Morgan of Princeton University described more precisely the nature of the agent responsible for Hawaiian volcanism. He hypothesized that hot spots are surface manifestations of narrow plumes of unusually warm material that rises through the earth's mantle. Although the mantle rocks are nominally solid, they exist under such high temperatures and intense pressures that they flow slowly, forming giant currents that rise and fall within the earth. In this scenario, hot spots form above an upwelling current. The current should produce not only a locus of volcanic activity but also a broader, upward bulge in the surface around the hot spot. Measurements of the shape of the earth show that the seafloor extending several hundred kilometers around the big island of Hawaii is indeed elevated as much as a few kilometers above the surrounding crust.

The mantle plume model does not identify where the hot-spot magma originates. Analyses of the mineral composition of Hawaiian lavas indicate that they fall into at least two chemically distinct categories. Presumably, one kind of lava derives from the same part of the upper mantle that supplies basaltic rock for new seafloor at mid-ocean ridges, where plates pull apart and hot rock rises from below. (The best-known feature of this kind is the Mid-Atlantic Ridge, which runs above sea level across Iceland.) Little is known about the source of the other lava component. Many geochemists suspect that it consists of minerals from a deep region of the mantle that still retains the chemical makeup it had when the earth formed.

As solid material ascends from the deep, it undergoes a tremendous decrease in pressure. The reduced pressure enables some previously solid minerals to melt. Because of its lower density, the lighter, melted fraction of the rock moves upward, separating from the solid residue as it rises. (The molten rock is referred to as magma while underground and as lava when it reaches the surface.) Laboratory experiments indicate that the rising magma collects in pockets. Those pockets of liquid rock force their way through the solid upper mantle and crust by creating and flowing through lens-shaped cracks. When the magma reaches the top of the crust, it provides the heat and raw material for Kilauea's eruptions.

Obtaining a more detailed picture of the workings of a volcano necessitates intimate acquaintance with one of these infernal creatures. To that end, Thomas A. Jaggar, Jr., who helped to pioneer the study of active volcanoes, founded the Hawaiian Volcano Observatory in 1912. The observatory is located on Uwekahuna ("bluff of the wailing priests"), a 100-meter-high cliff, where ancient priests once stood watch over Halemaumau. The priests who witnessed the antics of the fiery goddess Pele noted that the floor of Halemaumau often dropped abruptly at the same time as the red glow from an eruption appeared along a flank. This observation suggested that Pele moved along a subterranean route from Halemaumau to lower reaches of Kilauea. From that same vantage point at Uwekahuna, the three of us and our many colleagues, past and present, have watched Kilauea unleash its fury.

The surface features of the volcano are quite complex. Its summit gently rises more than one kilometer above sea level and more than six kilometers above the seafloor. Halemaumau forms the central part of a broad, shallow collapsed region, called a caldera, which extends as much as five kilometers across in places. Two long, elevated ridges, or rift zones, stretch away from the summit; they somewhat resemble the mid-ocean ridges in structure. The rift zone extending to the southwest is comparatively quiescent. Most of the recent volcanic activity at Kilauea has taken place along the other rift, to the east. That zone runs more than 100 kilometers from the summit, far out into the Pacific Ocean. Both rifts are surrounded by many parallel faults, cracks and cones of rock expelled during past eruptions.

As one learns to interpret its language, Kilauea begins to speak eloquently, through its ever shifting appearance, about the unseen processes taking place below. In the 1910s Jaggar noted that the volcano's summit region undulates repeatedly, rising slowly and then subsiding rapidly, often coincident with an eruption along a rift zone. He suggested that those movements are caused by the slow accumulation of magma in a shallow reservoir beneath the summit, followed by its rapid draining when the volcano erupts. In Jaggar's model the reservoir acts like a balloon buried within the volcano. The surface expands and bows upward when the reservoir inflates with magma. Subsequently, the surface subsides when magma escapes aboveground and the reservoir deflates.

Theoretical models predict that a small balloonlike pressurized cavity embedded in the earth would produce a dome-shaped uplifted area at the surface. Measurements of the uplift pattern in the summit region of Kilauea and several other volcanoes fit the models surprisingly well. In 1958 Kiyoo Mogi of the Earthquake Research Institute in Tokyo used a theoretical model of a pressurized cavity and data on the surface movements of Kilauea to estimate that the magma reservoir lies only three to four kilometers underneath the summit, well within the volume of the volcano.

During the late 1960s, Richard Fiske and Willie Kinoshita, who are both at the Hawaiian Volcano Observatory, confirmed Mogi's depth estimate and found that the center of the uplift occasionally migrates within the summit area. Such behavior implies the reservoir consists not of a single volume but of connected chambers that fill with magma at variable rates.

In the past, geophysicists relied on standard land-surveying techniques to measure Kilauea's constantly changing topography. That approach requires repeated measurements of hundreds of reference points on the volcano's surface. Within the past few years, an array of earth-orbiting satellites, called the Global Positioning System (GPS), has come into operation. Radio transmissions from the satellites are picked up by receivers distributed across the volcano that read out their relative positions to within a few centimeters. As magma flows beneath the volcano, the ground surface slightly rises or sags, changing the positions of the surveyed points. We have used GPS to great effect in following subtle changes in the volcano's shape, which in turn reveal the depth and volume of magma.

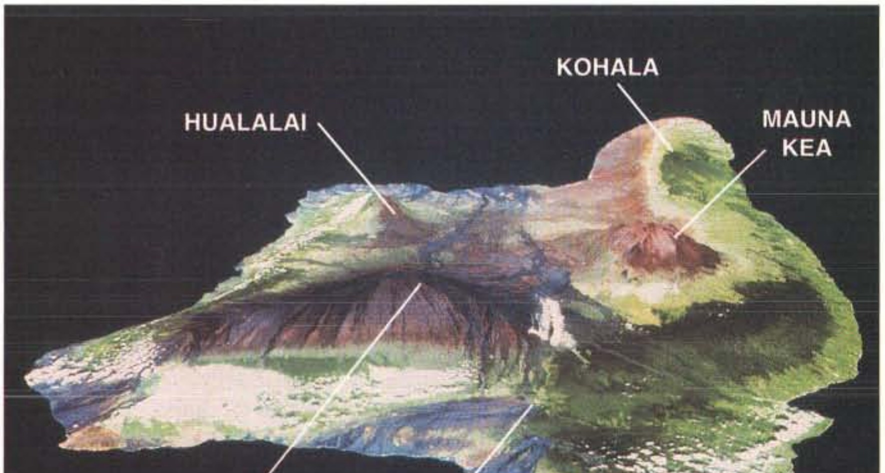
Researchers have recently begun probing the insides of Kilauea using tomography, a powerful method of analysis developed for medical imaging. In this approach, investigators examine seismic waves that have passed through the volcano. By studying changes in the speed and direction of the waves that take place during their journey, it is possible to infer internal structure. Because the earth around Kilauea is in a stressed state, there is no shortage of seismic signals: an earthquake 4.0 or higher in magnitude, sufficient to rattle dishes, occurs roughly every month. Late in the 1980s Phyllis Ho-Liu, then at the California Institute of Technology, produced tomographic images of the magma reservoir under Kilauea's summit. Her maps show that the reservoir lies within a few kilometers of the surface.

It may seem odd that after its long journey through the earth's mantle into the crust, rising magma would pause to accumulate in a reservoir so close to the surface. Michael Ryan of the U.S. Geological Survey in Reston, Va., attributes this behavior to the slight difference in density between the rising magma and the rock that forms the uppermost part of the volcano. As magma rises to the surface, dissolved gases come out of solution (like carbon dioxide in an opened bottle of soda water), filling the rock with bubbles. Because of these gas-filled spaces, lava that has erupted is 10 to 20 percent less dense than it was when it lay underground; in a sense, the volcano floats atop the magma reservoir.

Over time, lava from Kilauea is buried under flows from subsequent eruptions. At some depth below the volcano's surface, the weight of those overlying flows crushes and closes the bubbles. Lava therefore grows ever more compressed as depth increases, until its density matches that of the magma. At Kilauea, that equilibrium occurs about three kilometers below the surface, thus explaining the depth of the magma reservoir. Magma rises above this level only if impelled by additional pressure, which may be exerted by a deeper magma source derived from the hot-spot plume.

Several lines of evidence imply that all lava at Kilauea passes through the summit reservoir. The summit always subsides at the time of rift eruptions, demonstrating that magma is escaping from the reservoir. Studies at the Hawaiian Volcano Observatory show that the volume of subsidence equals or exceeds the volume of lava that emerges along the rift, except when an eruption lasts more than about one month. During extended events, magma probably flows quickly through the entire conduit system, rising in one continuous motion from the upper mantle through the summit reservoir, into the rift zone and onto the surface.

Chemical variations in lavas from Kilauea also point to the central role of the summit reservoir. The composition of lava changes during some rift eruptions of long duration, such as those that took place in 1955 and 1983. Lavas that emerge at the beginning of a rift eruption contain minerals that crystallize at relatively low temperatures. These comparatively cool lavas most likely underwent extended storage un-

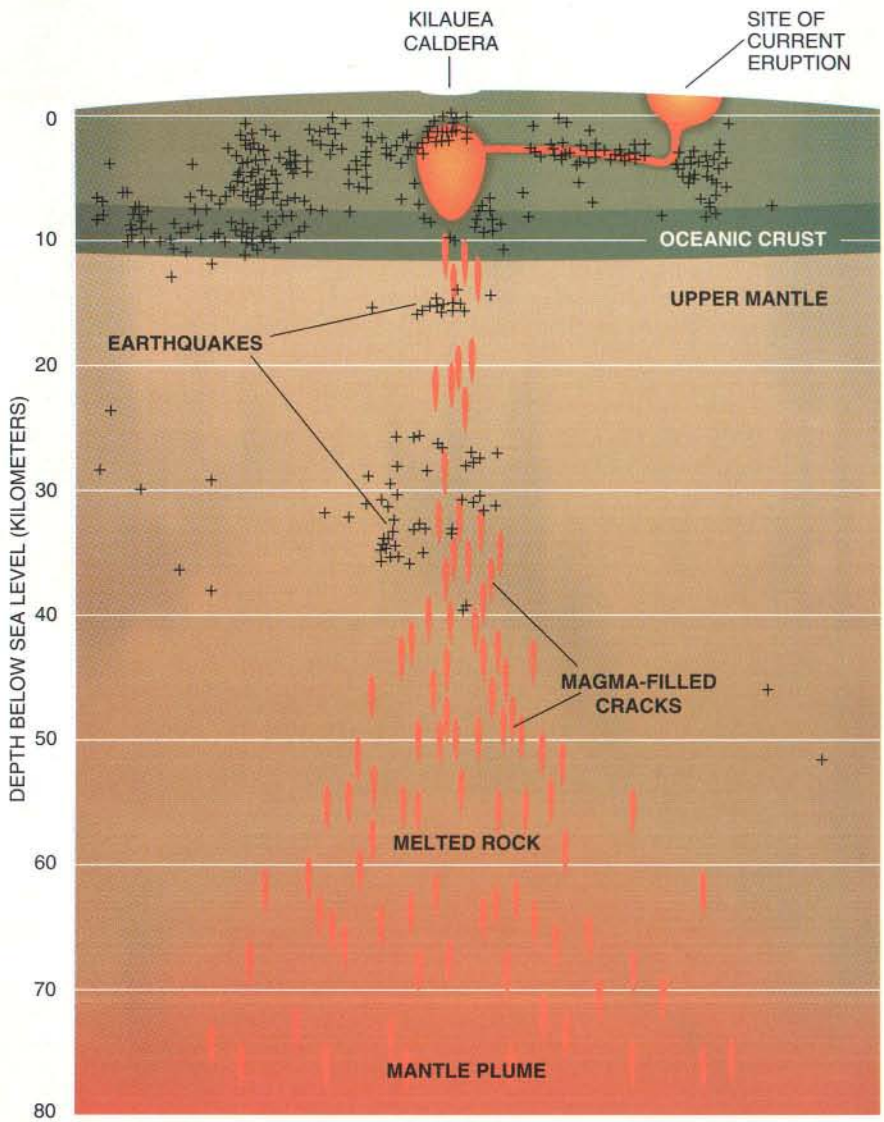


LARGE ISLAND OF HAWAII consists of five volcanoes, but only Kilauea is currently erupting. Kohala is considered extinct. Mauna Kea and Hualalai have lain dormant for 4,000 years and 190 years, respectively. Mauna Loa last erupted in 1984.

derground at a shallow level, presumably in the summit reservoir or in a secondary reservoir within the rift zone, where the temperature of the molten rock dropped sufficiently for many kinds of minerals to precipitate.

As a rift erupts, the lava stored in the summit reservoir should gradually be flushed out by fresh, hot magma rising from below. Sure enough, lavas that surface later in a rift eruption are richer in olivine, the first mineral to crystallize as magma cools. Many of the low-temperature minerals seen in the earlier lavas are notably absent. The lavas that erupt later emerge having a higher temperature than do their predecessors, a sign that the rock has spent little time cooling at a shallow level.

Additional evidence that the summit reservoir functions as a lava clearinghouse comes from observations made during brief lulls in the volcanic activity along Kilauea's east rift zone. At the beginning of such respites, swarms of shallow earthquakes occur along the east rift zone between the summit and the eruptive vent. These earthquakes probably occur along the top of a horizontal conduit that transports magma from the summit reservoir to the eruptive vent. Temporary restrictions in the flow path cause pressure to build up at the summit end of the conduit, triggering the earthquakes. Ho-Liu's tomographic maps confirm the presence of a horizontal conduit running from the summit reservoir to the east rift zone; the conduit intersects a magma reservoir under the present eruptive vent and ends under a prehistoric, now dormant vent called Heiheiahulu.



KILAUEA VOLCANISM is caused by a rising column of hot material that partially melts as it decompresses. Buoyant magma forces its way through the surrounding rock, generating earthquakes (*black crosses*) as it moves. Magma collects in a reservoir a few kilometers below Kilauea's summit; the molten rock may erupt upward or flow through a horizontal conduit to emerge along the volcano's flank.

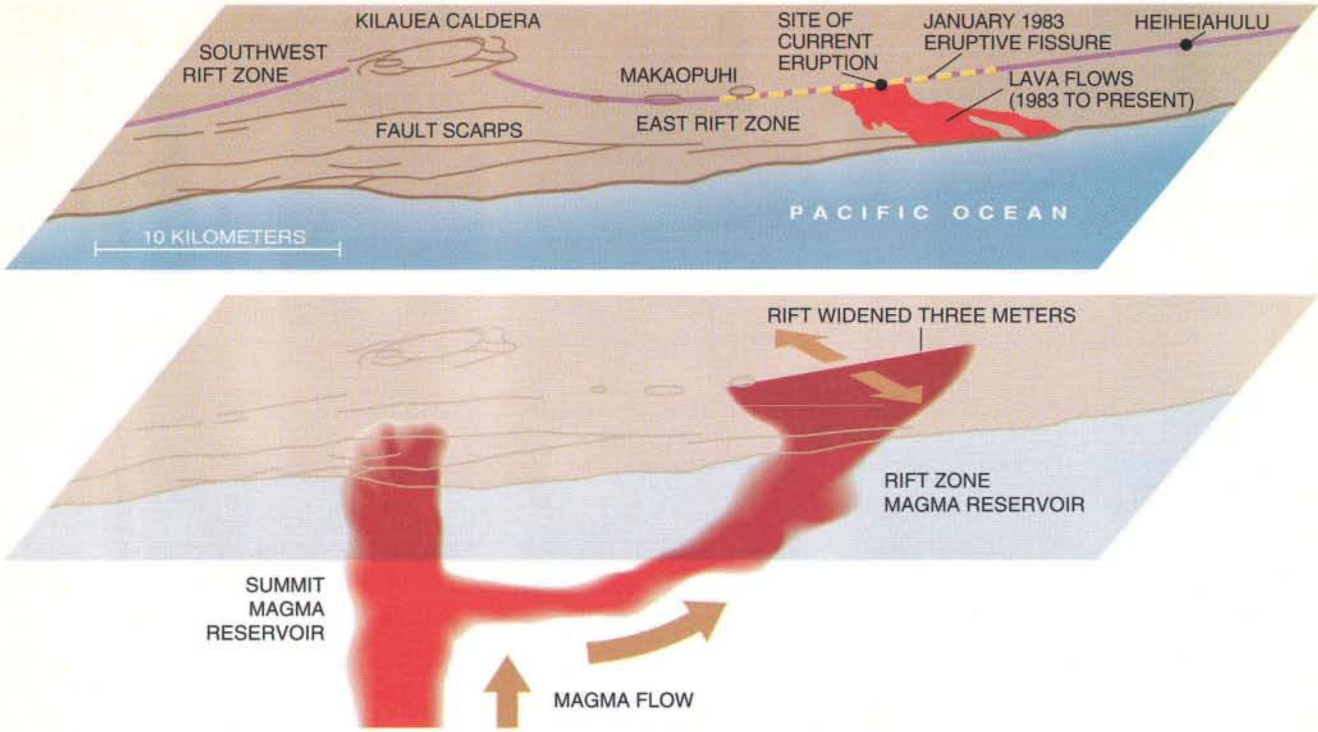
The accumulation of magma and the related increase of pressure within the volcano ultimately cause all of Kilauea's activity. Hence, calculating the magma supply rate and discovering how it varies are fundamental to understanding the dynamics of the volcano. Because all lava at Kilauea seems to have passed through the summit reservoir, it should be possible to infer the rate at which molten rock rises from the mantle by measuring how quickly magma flows into and out of the summit reservoir. This calculation requires knowing the pace at which magma escapes from the reservoir, along with any net change in the volume of magma stored there.

The amount of lava erupted is probably somewhat less than the volume of magma flowing into the summit reservoir because magma exiting the reservoir may become trapped in underground compartments. But over a decade or more, the eruptive rate at Kilauea averages out to a constant fraction of the magma supply rate.

Researchers began systematically measuring changes in the volume of the summit reservoir in 1958. Since then, the volume of magma stored below the summit has fluctuated by only 0.5 cubic kilometer, judging from changes in the shape of Kilauea's surface. Over the same time span, some two cubic kilometers of lava issued from the volcano, showing that the change in the size of the reservoir has been small compared with the total volume of magma that has passed through it. Variations in the eruptive rate, if averaged over many years, should therefore reflect fluctuations in the rate at which magma is supplied from below.

From 1840 to the present, the eruptive rate at Kilauea, averaged over decade intervals, has varied from zero to 0.1 cubic kilometer per year. In the spring of 1950, after 16 years of no volcanic activity, the rate of earthquakes under the summit began to increase, and the area around the summit started to rise, suggesting that the magma supply rate had begun to increase. Since then, the volcano's eruptive rate has been considerably above its long-term average. The highest magma supply rate probably occurs when the volcano is in a continuous, or nearly continuous, state of eruption.

Donald Swanson of the U.S. Geological Survey in Seattle calculates that during three long eruptions, Kilauea produced about 0.1 cubic kilometer of lava per year, sufficient to fill a stadium the size of the Houston Astrodome in



MAGMA ACCUMULATES in connected reservoirs below Kilauea's summit caldera and under the east rift zone (*bottom*); eruptions occur when the magma pressure reaches a critical point. Since 1983, magma from the secondary reservoir has been erupting along a 12-kilometer segment of the east rift, giving rise to lava flows that extend to the Pacific Ocean (*top*).

one week. Swanson notes that the summit did not noticeably swell or subside during these eruptions, so the volume of magma in the reservoir must have remained nearly unchanged. During lengthy eruptions, magma evidently flows into the summit reservoir as fast as lava flows onto the surface. A consideration of the nature of rift eruptions shows why this should be so.

The movement of magma into and out of the summit reservoir is driven by a difference in pressure between the material in the reservoir and other pools of magma lying either deep under the summit or along the horizontal conduit below a rift zone. Along a rift zone the slope of the volcano primarily determines the magma's movement. The summit reservoir, located immediately below Kilauea's summit, resides under a much thicker layer of rock than do the comparatively shallow magma chambers under the rift zones. The great weight of the summit rocks tends to squeeze the magma into the rift zone.

A simple set of observations made at the Hawaiian Volcano Observatory has confirmed the effect of Kilauea's topography on eruptions. The volume of lava released during a rift eruption (and, by inference, the volume of magma withdrawn from the summit reservoir) is proportional to the elevation of the rift at which the eruption occurs. The more voluminous rift eruptions occur at lower elevations, where the difference in pressure between the summit reservoir and the eruptive vent is greater.

During a rift eruption, lava gushes rapidly aboveground, and magma drains from the summit reservoir suddenly, decreasing the pressure there. Consequently, a significant difference in pressure develops between the summit reservoir and the deeper magma sources that supply it. That difference in pressure causes magma to refill the summit reservoir rapidly, a process that manifests itself at the surface as a sudden uplift of the summit. As the shallow reservoir refills, the difference in pressure decreases, and the supply rate gradually slows. In this way, an equilibrium develops, as long as the magma supply rate is sufficient to maintain the prodigious flow of lava from the rifts.

Assuming the above reasoning is correct, one would expect that frequent rift eruptions would cause a high magma supply rate. Historical records show just such a pattern: only six rift eruptions took place between 1840 and 1950, whereas 17 rift eruptions have occurred since 1950, when Kilauea has spewed lava at a much greater rate. Pressure of the summit reservoir seems to control both the frequency of eruptions and the rate at which magma is supplied to the base of the volcano.

After an eruption ends, plugs of solidified rock block the path through which the magma traveled. Although the volcano may look peaceful from above, intense pressure builds below as magma continues to ascend from the mantle into the summit reservoir. When the reservoir again fills past its capacity, the surrounding rocks split, and new underground pathways form. Eventually magma bursts onto the surface, and another eruption, in a different location, is under way. This process unfolded right before our eyes, beginning in the fall of 1982.

A small eruption lasting a few hours on September 25 led the staff at the Hawaiian Volcano Observatory to suspect that the magma reservoir under Kilauea was filling beyond its capacity. A steady increase in elevation of the summit and an unusual abundance of shallow earthquakes centered below the summit bolstered that suspicion. The shaking of the ground signaled that the pressure in the summit reservoir had become high enough to crack the surrounding rocks.

During the week after the September eruption, the shallow earthquakes spread from the summit region to the adjoining segment of the east rift zone. These tremors traced the migration of magma through a subterranean con-

duit running from the summit reservoir out along that rift zone. Under the burgeoning pressure, rock around the conduit began to split and to shift, setting off the seismometers.

By December the continued accumulation of magma in the summit reservoir forced the liquid rock into the east rift zone toward Makaopuhi, a crater some 20 kilometers east of Kilauea's caldera. The Makaopuhi area was the site of several short-lived eruptions between 1963 and 1969. Researchers measuring ground motions at Kilauea deduced the presence of a secondary magma reservoir under Makaopuhi; some magma may have remained trapped in that reservoir.

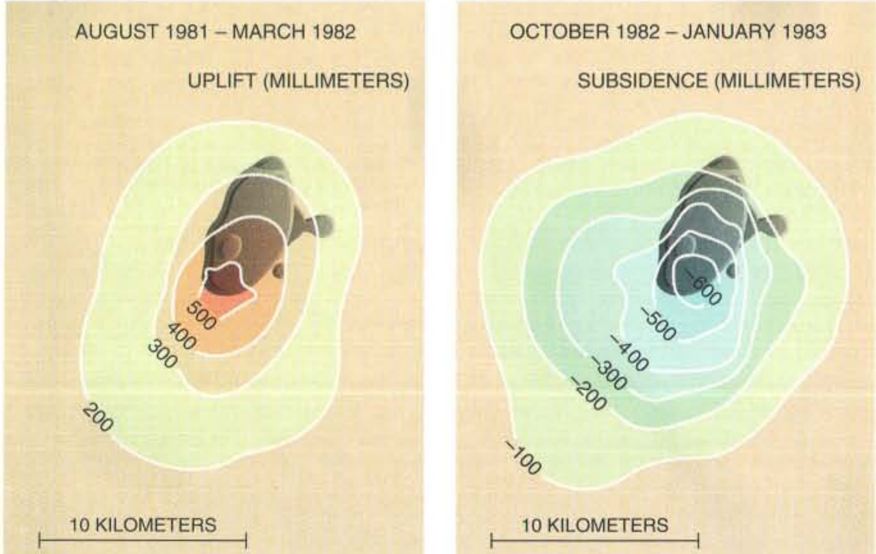
The earthquake record tells that the magma advanced from the summit reservoir in several short bursts, each lasting a few hours and extending a few kilometers farther from the caldera than did its predecessor. Finally, on January 2, 1983, a rapid succession of shallow earthquakes near Makaopuhi and a sudden uplift of the ground surface announced that magma had filled the secondary reservoir and begun pushing toward the surface.

KILAUEA SUMMIT CALDERA appears as an oval depression filling the bottom half of this 1989 photograph; the large circular crater is Halemaumau. The plume of smoke in the background identifies the ongoing eruption along the east rift zone.

Judging from the shape of the uplift, magma began its upward journey about three kilometers under Makaopuhi, probably at the top of a secondary magma reservoir. Computer simulations conducted in the laboratory reveal the nature of the magma's final ascent. Magma forcing its way into a tiny break at the top of the reservoir rapidly expanded horizontally and vertically while remaining very thin, so that it came to resemble a sheet. Lava spurted from the ground along a 12-kilometer-long fissure where the magma sheet intersected the surface. The formation of this fissure widened the east rift zone by about three meters, thereby compressing the adjacent part of the southern flank of Kilauea.

For the next three and a half years, Kilauea episodically spouted spectacular fountains of lava along its east rift zone. Most of the spurts lasted less than a day, followed by month-long quiescent periods. In the summer of 1986, after several new fissures opened, the lava fountains ceased, and since then, molten material has flowed fairly continuously into a large lava lake. At present, that lake has drained through lava tubes onto the surrounding landscape, and a new one has formed several kilometers closer to the summit caldera.

Clearly, living near an active volcano exposes one to some significant hazards. The molten rock and sulfurous fumes that accompany eruptions constitute the most obvious threat to human life. Violent steam explosions occasionally shake Kilauea; one in 1790 caused a number of fatalities. During the present eruption, lava has flowed into the ocean, unleashing a noxious cloud of steam laced with hydrochloric acid, which local residents call volcanic smog, or vog.



BULL'S-EYE VIEW of Kilauea's summit shows that the surface bowed upward (*orange colors*) while magma accumulated in the underlying reservoir (*left*). The summit subsided (*blue colors*) when lava started erupting onto the surface (*right*).

A less familiar natural hazard around Kilauea and many other volcanoes is the possibility of a major earthquake. The annual rate of earthquakes under Kilauea's southern flank is comparable to the frequency of earthquakes along the most active segments of the San Andreas fault in southern California and Mexico. Kilauea generates an earthquake that one can feel every two weeks or so. One of the most powerful earthquakes in the U.S. this century—magnitude 7.2 on the Richter scale—occurred below the volcano's southern flank on November 29, 1975. The Hawaiian Volcano Observatory has set up an islandwide network of four dozen seismometers to record earthquakes. An increase


TWO FACES OF KILAUEA seen in these photographs illustrate the mutable nature of the volcano. During the first three years of the present eruption, huge fountains of lava (*left*) periodically gushed from the east rift zone. Since 1986 the volcano has erupted more peacefully, creating a lava lake covered with a fractured crust of solidified rock (*right*).

in the frequency of shallow tremors (those centered no more than a few kilometers below the surface) has proved to be a good signal that the volcano is building up to an eruption.

Most earthquakes, such as those in California, result from the grinding movements of the earth's plates. The frequent earthquakes around Kilauea take place where rocks shift in response to changes in internal pressures. Earthquakes on Kilauea's southern flank result from bulk movement of magma, especially into a rift zone, as happened in January 1983. For several weeks after that event, the southern flank experienced an elevated rate of earthquakes, probably in response to sudden compression by the three-meter-wide opening of the rift zone. The continued above-average abundance of moderately strong (magnitude 5.0 or greater) earthquakes since 1983 results from rocks that are still reacting to the changed pressure.

To sharpen the understanding of that process of adjustment, the Hawaiian Volcano Observatory staff studied in detail the most powerful of the post-1983 tremors, a magnitude 6.1 earthquake that took place on June 25, 1989. We used the Global Positioning System to determine the exact locations of our measuring stations less than a year before and again two months after the earthquake. The latter measurements showed that the stations along the southern flank had moved seaward by as much as one third of a meter. Stations elsewhere on Kilauea also generally moved seaward, although by a lesser amount. Those shifts released some of the compressive pressure that had built up from the January 1983 eruption and from earlier rift events.

The insights gained by studying Kilauea are contributing to the general study of active volcanoes and of earthquakes. Although volcanoes vary widely in style of eruption and geologic setting, all active volcanoes seem to share a common trait: they are driven by magma ascending from deep regions in the mantle into a shallow reservoir in the earth's crust or within the volcano itself. Locating and mapping these reservoirs are central to the goal of predicting eruptions. The thoroughly scrutinized reservoir under Kilauea's summit has served as a starting model for understanding the structure of other active volcanoes.

Kilauea has also provided a valuable field experience for scientists who later have worked under the hectic and, at times, dangerous conditions that accompany the eruption of any volcano. The rapid scientific response to the reawakening of Mount St. Helens in 1980 drew on earthquake and ground movement monitoring techniques developed at Kilauea. Similarly, techniques developed at Kilauea—especially the study of earthquake patterns—were critical to the successful prediction of the climactic eruption of Mount Pinatubo in June 1991, the second or third most powerful volcanic event of this century. The warning saved thousands of lives and gave sufficient time to evacuate expensive equipment and aircraft from nearby Clark Air Force Base.

Given the earth's restless nature, it is inevitable that additional volcanic crises will occur. The potential for disaster continues to increase as populations expand to areas where there is a high risk of volcanism. Protection of these people will depend on the ongoing efforts to predict the behavior of volcanoes and to understand the geologic processes that endow them with their devastating fury.

FURTHER READING

HOW VOLCANOES GROW. J. P. Eaton and K. J. Murata in *Science*, Vol. 132, No. 3432, pages 925–938; October 7, 1960.

CHEMICAL VARIATION IN KILAUEA ERUPTIONS 1971–1974. Thomas L. Wright and Robert I. Tilling in *American Journal of Science*, Vol. 280-A: *The Jackson Volume*, Part 2, pages 777–793; 1980.

VOLCANISM IN HAWAII. Edited by Robert W. Decker, Thomas L. Wright and Peter H. Stauffer. U.S. Geological Survey Professional Paper 1350. U.S. Government Printing Office, 1987.

NATURE OF LOCAL MAGMA STORAGE ZONES AND GEOMETRY OF CONDUIT SYSTEMS BELOW BASALTIC ERUPTION SITES: PU'U 'O'O, KILAUEA EAST RIFT, HAWAII, EXAMPLE. Lionel Wilson and James W. Head III in *Journal of Geophysical Research*, Vol. 93, No. B12, pages 14,785–14,792; December 10, 1988.

Bacterial Endotoxins

An integral part of many bacteria, these molecules are at once brutal and beneficial to humans. Efforts are under way to block the bad effects and harness the good

by Ernst Theodor Rietschel and Helmut Brade

Endotoxins possess an intrinsic fascination that is nothing less than fabulous. They seem to have been endowed by nature with virtues and vices in the exact and glamorous proportions needed to render them irresistible to any investigator who comes to know them.

This observation, made by Ivan L. Bennett, Jr., of Johns Hopkins University almost 30 years ago, is still true today. Endotoxins, which are potentially lethal molecules produced by many bacteria—including those responsible for cholera, whooping cough, plague and a form of meningitis—continue to captivate medical researchers and others.

Of the vices that render endotoxins "irresistible," perhaps the most impressive is the power to elicit disease symptoms ranging from chills and fever to irreversible shock (circulatory failure leading to malfunctioning of organs throughout the body) and death. The molecules also form a rigid shield on endotoxin-bearing bacteria, preventing many antibiotics from combating infection by those bacteria.

ERNST THEODOR RIETSCHEL and HELMUT BRADE are colleagues at the Borstel Research Institute in Borstel, Germany. Rietschel, who is director of the institute, has studied endotoxins for more than 20 years. He received his Ph.D. in 1971 from the University of Freiburg, while he was working with Otto Westphal and Otto Lüderitz at the Max Planck Institute for Immunology in Freiburg. In 1980 he became professor of immunochemistry and biochemical microbiology at the Medical University of Lübeck and also assumed his current position at the Borstel institute. Rietschel has been president of the International Endotoxin Society since 1990. Brade earned an M.D. from the University of Düsseldorf in 1976 and was a research fellow under Lüderitz in Freiburg before joining Rietschel's laboratory in 1983. He is now associate professor and head of the division of biochemical microbiology at the Borstel institute.

Paradoxically, the same endotoxins that threaten human health can enhance the body's overall immune resistance to bacterial and viral infections and cancer. Some investigators think exposure to these molecules may even be essential to the development and ongoing viability of the immune system.

As a crucial step toward understanding how endotoxins might achieve such diverse and contrary effects, dozens of research teams, including our group at the Borstel Research Institute in Germany, have concentrated on elucidating their chemical and three-dimensional structure. The findings have helped identify the components responsible for the activities of endotoxin in the body. The work has also begun to generate ideas both for limiting the ability of these substances to cause disease and for harnessing their remarkable tendency to heighten one's disease-fighting capacity.

Endotoxin's story rightly begins in the late 1870s, soon after Robert Koch, then working on his own in Wollstein, Germany, established that each infectious disease stems from a specific microbe. Within a decade after Koch's demonstration, investigators in France, Germany, the U.S. and elsewhere had shown that bacteria often make people sick by producing toxins.

The first poisons isolated were exotoxins, substances secreted actively by many bacteria, such as those responsible for diphtheria, tetanus and botulism. Today researchers know that exotoxins are generally proteins, but at the time, a major identifying feature was susceptibility to inactivation by heat.

Then, in 1892, two scientists separately described poisons that did not fit the exotoxin profile. Richard Pfeiffer, one of Koch's students, found that the *Vibrio cholerae* bacterium, which causes cholera, synthesized not only a heat-labile exotoxin but also a heat-resistant substance that was not secreted by the bacterial cell. (Bacteria are single-cell organisms.) This second product seemed to be released only when *V. cholerae* disintegrated. Heat-resistant toxins had been studied earlier, particularly by the Danish pathophysiologist Peter Ludwig Panum, but only Pfeiffer noticed they were not actively extruded.

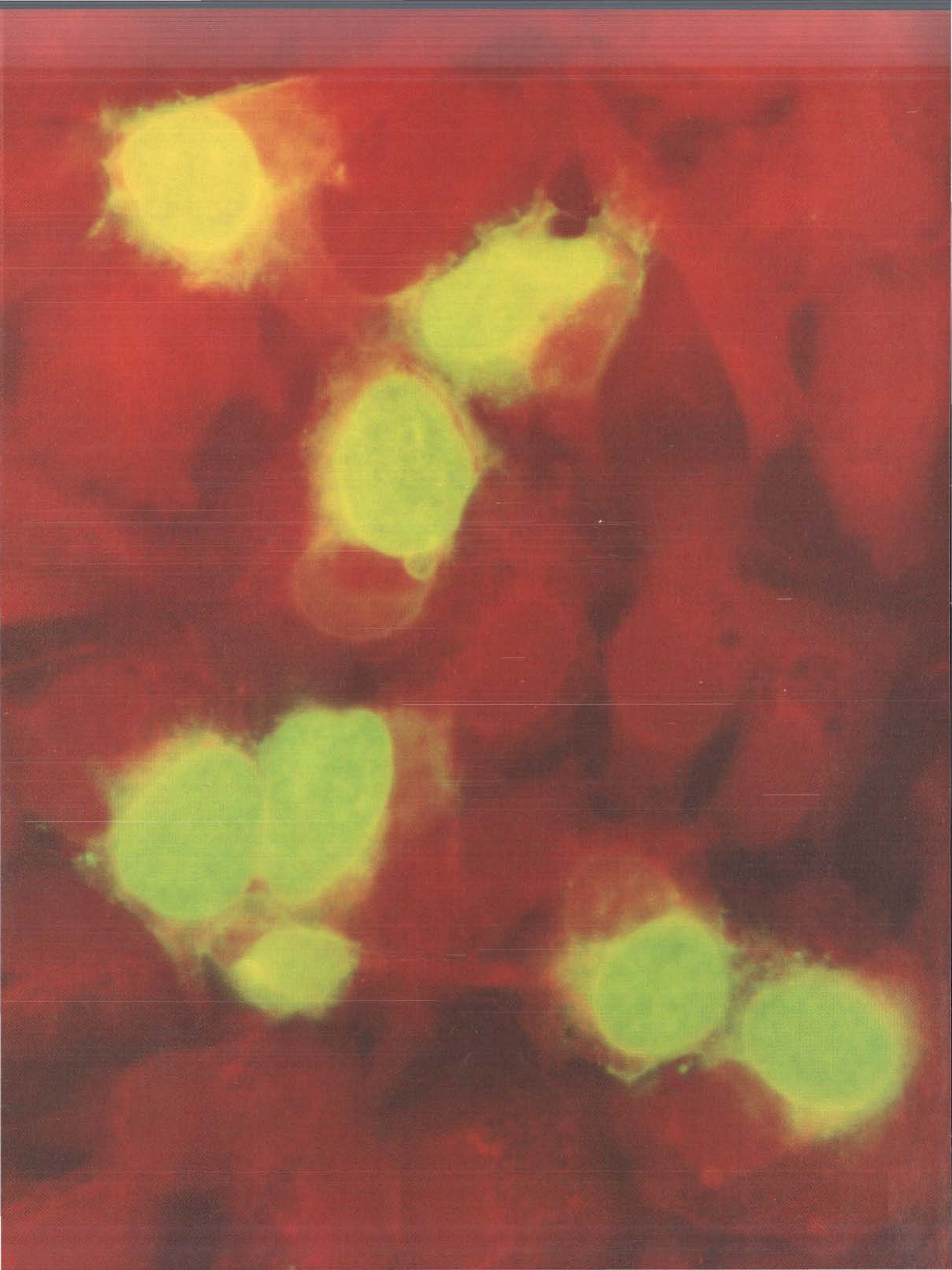
Assuming the second substance was sequestered within bacteria, Pfeiffer added the Greek word for "within" to "toxin" and named his discovery endotoxin. Subsequent work has revealed the term to be a misnomer: endotoxins reside on the surface of bacteria, not in the interior [*see illustration on page 28*].

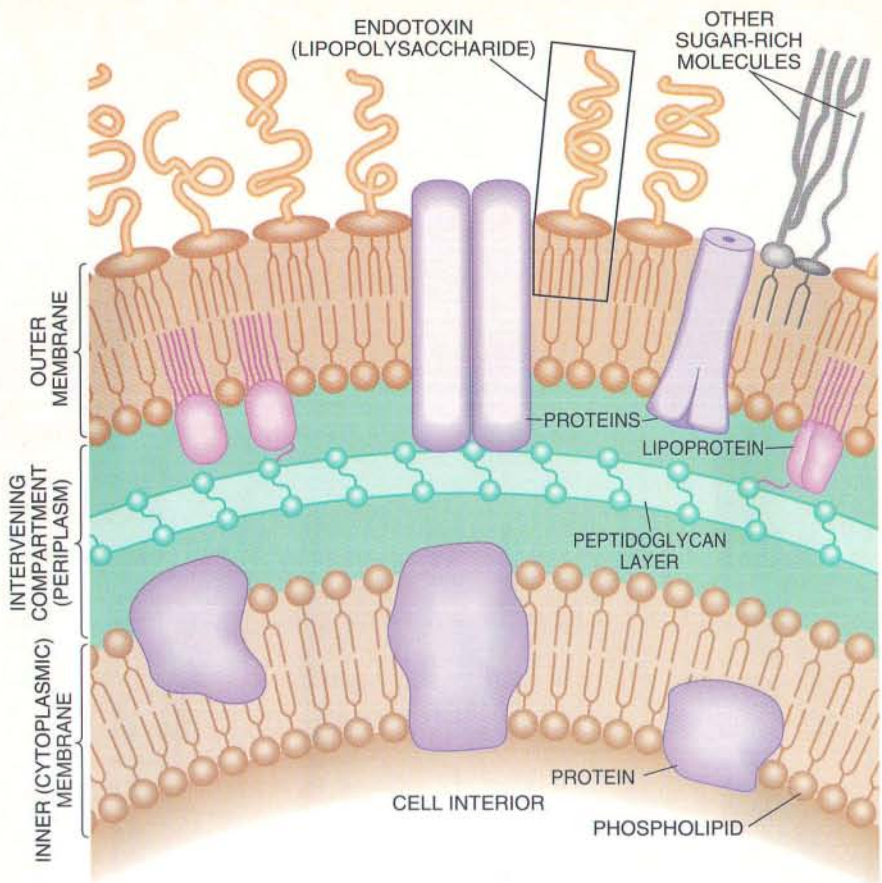
While Pfeiffer toiled in Germany, Eugenio Centanni of the Institute of Pathology at the University of Bologna extracted a heat-stable toxin from *Salmonella typhi*, the bacterium that gives rise to typhoid fever. Because this poison caused fever in rabbits, Centanni called it pyrotoxina.

Centanni's pyrotoxina and Pfeiffer's endotoxin are now known to be basically identical. It is also evident that such toxins are characteristic of all so-called gram-negative bacteria. (Bacteria are classified as gram-positive or gram-negative depending on whether or not, in a procedure devised by the Danish physician Hans Christian Joachim Gram, they retain a particular blue dye.)

Such realizations awaited advances in chemical technology. André Boivin of the Pasteur Institute in Paris, Walter T. J. Morgan of the Lister Institute in London and Walther F. Goebel of the Rockefeller Institute in New York City

ENDOTOXIN MOLECULES made by *Chlamydia trachomatis* bacteria that have multiplied within mouse fibroblasts (*red spheres*) are highlighted by fluorescence (*yellow-green*). Endotoxins account for many of the disease symptoms caused by gram-negative bacteria, a category that includes organisms responsible for such disorders as cholera, whooping cough, the plague, and certain forms of food poisoning, pneumonia and meningitis. The fibroblasts in the image are enlarged some 1,500 diameters.





ENDOTOXINS RESIDE in the outer membrane of the bacterium *Escherichia coli* (*micrograph*), a cross section of which is shown schematically. The cut reveals the structure of the two membranes that envelop the bacterium. An endotoxin-rich outer membrane is an identifying feature of gram-negative bacteria.

made the first major analytic strides in the 1930s and 1940s. Working with relatively impure extracts—the best that could be obtained at the time—the three concluded that the heat-stable toxin in every gram-negative bacterium they examined contained polysaccharide (polymerized sugars), lipid (complexes containing fatty acids) and protein.

In the 1940s as well, Murray J. Shear, then at the National Institutes of Health, analyzed the active component of the gram-negative bacterium *Serratia marcescens,* which can cause nosocomial, or hospital-acquired, infections. He was intrigued by this component because William B. Coley of Memorial Hospital in New York City had shown that mixtures of killed *S. marcescens* and *Streptococcus* could sometimes lead to regression of cancer. In 1943 Shear determined that the toxic and tumor-destroying material in *S. marcescens* consisted mainly of polysaccharide and lipid. Emphasizing chemistry, he called it lipopolysaccharide, or LPS.

By then the accumulating evidence was already suggesting that lipopolysaccharide, endotoxin and pyrotoxina were essentially the same. But convincing proof was obtained only after Otto Westphal and Otto Lüderitz developed a technique for generating large quantities of very pure extract from bacteria. They achieved this breakthrough in the late 1940s at the Wander Research Institute in Säckingen, Germany, which later became the Max Planck Institute for Immunology in Freiburg.

After applying the method to *Salmonella, Serratia, Vibrio* and related bacteria, Westphal and Lüderitz established conclusively that the toxic component always consists only of polysaccharide and lipid, along with phosphorus (which influences the behavior of the two other components). The protein found in earlier extracts was not needed for toxicity.

Further studies revealed that, regardless of origin, each isolated toxin led to the same responses in animals. This result implied that the molecules were, indeed, virtually alike chemically. Since then, the term "pyrotoxina" has fallen by the wayside, but "endotoxin" and "lipopolysaccharide" are applied interchangeably. Westphal and Lüderitz also helped to demonstrate that all gram-negative bacteria, and only those organisms, make endotoxins.

Establishing that all endotoxins are variations of a single molecule was a major accomplishment, but in order to understand how they work, further details of chemistry had to be deciphered. Because the molecules are so complex, investigators needed some 25 years more to arrive at a conclusive picture. Not surprisingly, the effort involved an extraordinary number of researchers—too many to mention individually here.

Most of the studies focused on the broad group of gastrointestinal (enteric) bacteria that form part of the normal flora of the intestine or cause gastrointestinal disease. Among these are members of the *Salmonella* and *Escherichia* genera, of which *E. coli,* a ubiquitous tool of molecular biologists, is a member. Most strains of *E. coli,* such as those commonly found in the intestine, are benign (at least when they remain in the gut), but others may lead to severe disease.

In the course of such studies, workers learned that the lipid component, called lipid A, is embedded in the outer membrane, or "skin," of bacterial cells, which in contrast to human cells have no nucleus. In fact, lipid A constitutes much of the external half of the membrane, which has two layers. The polysaccharide protrudes into the environment like hair on a head. Extending from the lipid, it is composed of two distinct parts: the core oligosaccharide (a small string of sugars), which is connected to the lipid, and the longer O-specific chain, which projects from the core and is the outermost part of the endotoxin [*see box on opposite page*].

The O-specific chain turns out to be the most variable segment and the part that evokes a specific immune reaction; that is, it stimulates production of antibody molecules able to recognize that particular O chain but no other molecules. In the 1960s Westphal and Lüderitz made the fundamental discovery that the chain usually differs from one bacterial species to the next. Hence, each variant of, say, *Salmonella,* will sport a distinctive O chain and elicit

production of different antibodies. Typically the chain consists of 20 to 40 repeating units that include up to eight sugars. The types and sequence of sugars within the basic unit and the number of repeating units often differ.

The core oligosaccharide, divided into the inner core (the part linked to the lipid) and the outer core (the part linked to the O-specific chain), is not so variable. Its influence on the body is also less profound, although it can trigger antibody production in response to mutant endotoxins that lack an O chain.

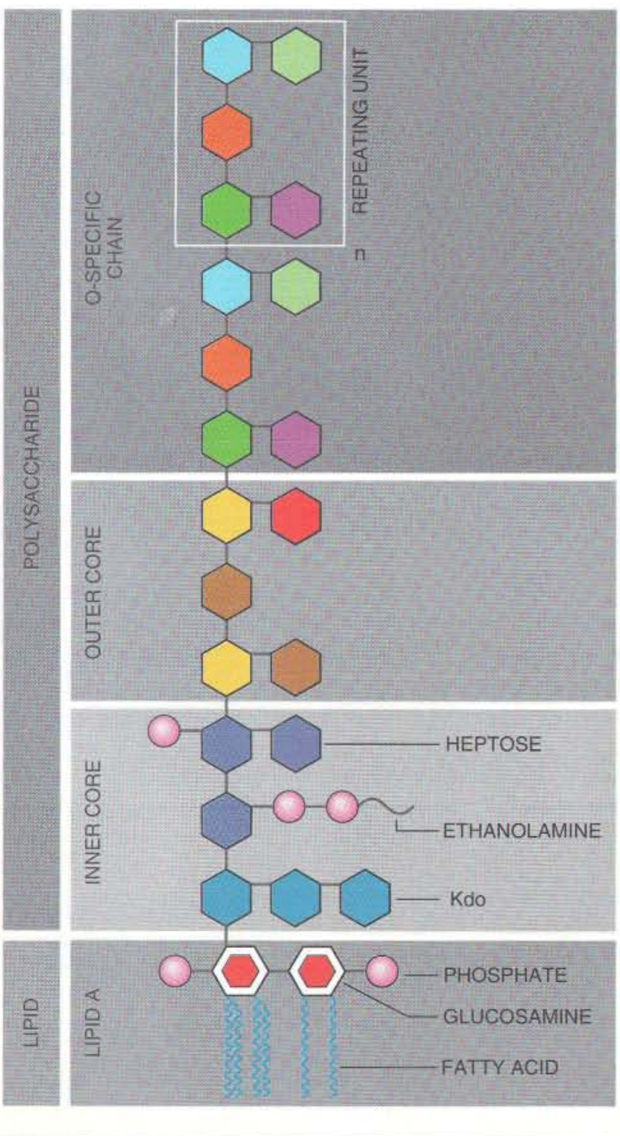
The inner segment is the more intriguing of the two because it bears unusual sugars. One is heptose, which boasts seven instead of the more typical six carbon atoms. The other is Kdo (3-deoxy-D-manno-2-octulosonic acid), which is found in all endotoxins and links the polysaccharide to the lipid. Kdo harbors eight carbon atoms and occurs nowhere else in nature, except in certain plants and algae. As will be seen, its ubiquity and uniqueness make it a potential target of therapy.

In terms of human health, the third component, lipid A, is the most fascinating. This region, which is the least variable in structure, gives rise to all the ills endotoxins inflict, as well as to the benefits—namely, increased resistance to infection and cancer.

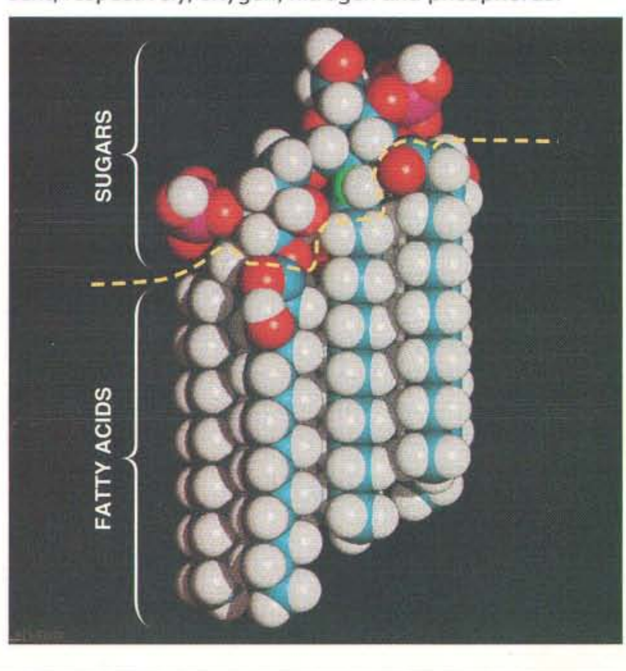
As long ago as 1954, Westphal and Lüderitz found that lipid A consisted mainly of a sugar—glucosamine—as well as phosphate (PO_4) and long chains of fatty acids, each with a backbone of about 14 carbon atoms. Yet how these constituents fit together remained for the most part unknown until 1969. Then Jobst Gmeiner, working in Westphal and Lüderitz's laboratories, showed that lipid A in one strain of *Salmonella minnesota* included two linked glucosamine sugars, each of which was phosphorylated (had a phosphate group attached).

Moreover, the carbon atoms linking the two glucosamines were joined in an unexpected way. The carbon atoms in sugars are numbered by standard conventions. In this case, the first carbon of one glucosamine was linked to the sixth carbon of the other glucosamine through an oxygen, so that the bridging oxygen was elevated above the glucosamines (in the so-called beta configuration). Most beta-linked glucosamine pairs are joined by oxygen bridges between carbons one and four, or some-

The Structure of Endotoxin

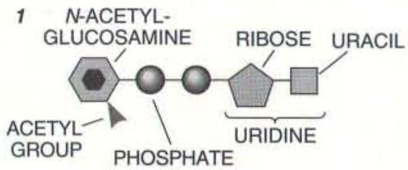


Chemical structure of the endotoxin made by *Salmonella typhimurium* (*left*), like that of other endotoxins, consists of a polysaccharide, or long chain of sugars (*each color represents a different sugar*), and a fat called lipid A. The polysaccharide, which varies from one bacterial species to another, is made up of the O-specific chain (built from repeating units of three to eight sugars) and the two-part core. Lipid A virtually always includes two glucosamine sugars modified by phosphate (PO_4) and a variable number of fatty acids. Manfred Kastowsky of the University of Berlin and Harald Labischinski of the Robert Koch Institute in Berlin have modeled the three-dimensional structure of lipid A in *Escherichia coli* (*below*), which is identical to the lipid in *S. typhimurium*. The fatty acid chains, made mainly of carbon (*blue*) and hydrogen (*white*), dangle from the sugars (*tilted region at top*). Red, green and pink spheres represent, respectively, oxygen, nitrogen and phosphorus.

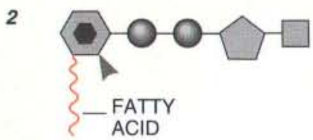


How Bacteria Synthesize Lipid A

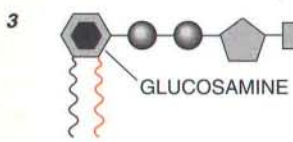
The synthesis of lipid A is depicted for *E. coli* but is similar for other gastrointestinal bacteria. (At each step, changes in the prior structure are indicated in red.) The first five steps were uncovered by Christian R. H. Raetz and his colleagues at the University of Wisconsin; the final steps by Mary Jane Osborn and her co-workers at the University of Connecticut Health Center.



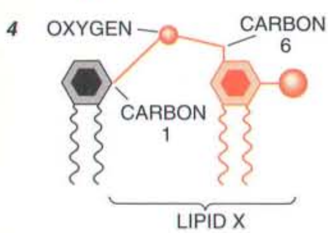
1. The bacterium begins constructing lipid A from a precursor consisting of the sugar *N*-acetylglucosamine bound through two phosphates to uridine.



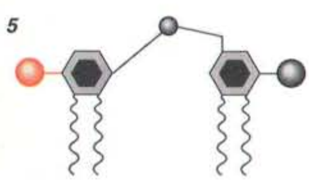
2. Then one fatty acid is added to *N*-acetylglucosamine.



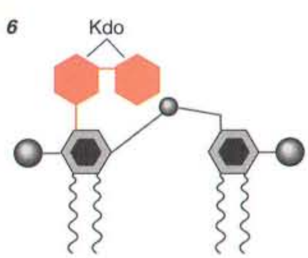
3. Removal of the acetyl group converts *N*-acetylglucosamine into glucosamine, which gains a fatty acid in place of the acetyl group.



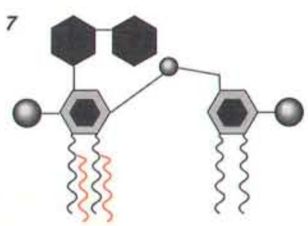
4. Next, an enzyme-controlled process replaces the two phosphates and uridine with lipid X, a glucosamine sugar carrying a phosphate group and two fatty acids.



5. Now the original glucosamine gains a phosphate group...



6. ... after which two Kdo sugars of the inner core, which link lipid A to the rest of the core, are added to the glucosamine.



7. Lipid A is completed with the addition of more fatty acids. (Then the balance of the core is attached to the Kdo bound to lipid A, and the O-specific chain is added.)

Interference with construction of lipid A prevents bacteria from producing viable offspring. Kdo analogues, which block growth of lipid A after step 5 and therefore halt development of the rest of the endotoxin molecule, are being studied as antibiotics.

times one and three. A one to six linkage was unheard of, however.

Later, one of us (Rietschel) and Sumihiro Hase, working with Westphal and Lüderitz, demonstrated that the same beta 1-6 linkage of glucosamines occurred in many gram-negative bacteria. Today it is evident that this arrangement is characteristic of, and unique to, most gram-negative bacteria.

Determining where the fatty acids bind to the glucosamines was rather more difficult. In fact, it took many researchers, among them Ulrich Zähringer, Ulrich Seydel and Rietschel of Borstel, until 1983 to decipher the complete structure of *E. coli* lipid A.

Simultaneously, Nilofer Qureshi and Kuni Takayama of the University of Wisconsin determined the organization of lipid in *Salmonella typhimurium* (a cause of diarrhea). It was identical to that of *E. coli*. In other bacteria, the length of the fatty acid chains and attachment sites can differ. At the least, though, endotoxins of disease-causing gram-negative bacteria typically include, along with phosphorylated glucosamines, four fatty acids that carry a hydroxyl (OH) group on the third carbon in each string, as well as one or two nonhydroxylated fatty acids. In some instances, glucosamine may be replaced by an unusual form, 3-amino glucosamine.

As the structural work was proceeding, investigators were collecting increasing evidence that lipid A was responsible for the harm done by endotoxins and for enhancement of immunity. Westphal and Lüderitz had suggested that probability as early as 1954, when it became apparent that the polysaccharide was too variable to elicit a consistent set of effects.

This deduction gained support in the late 1960s, when Yoon B. Kim and Dennis W. Watson of the University of Minnesota and other research teams established that endotoxins consisting only of lipid A and Kdo were as toxic and pyrogenic as molecules that included a full polysaccharide. That finding proved that the O-specific chain and large parts of the core were not needed for producing toxicity and fever.

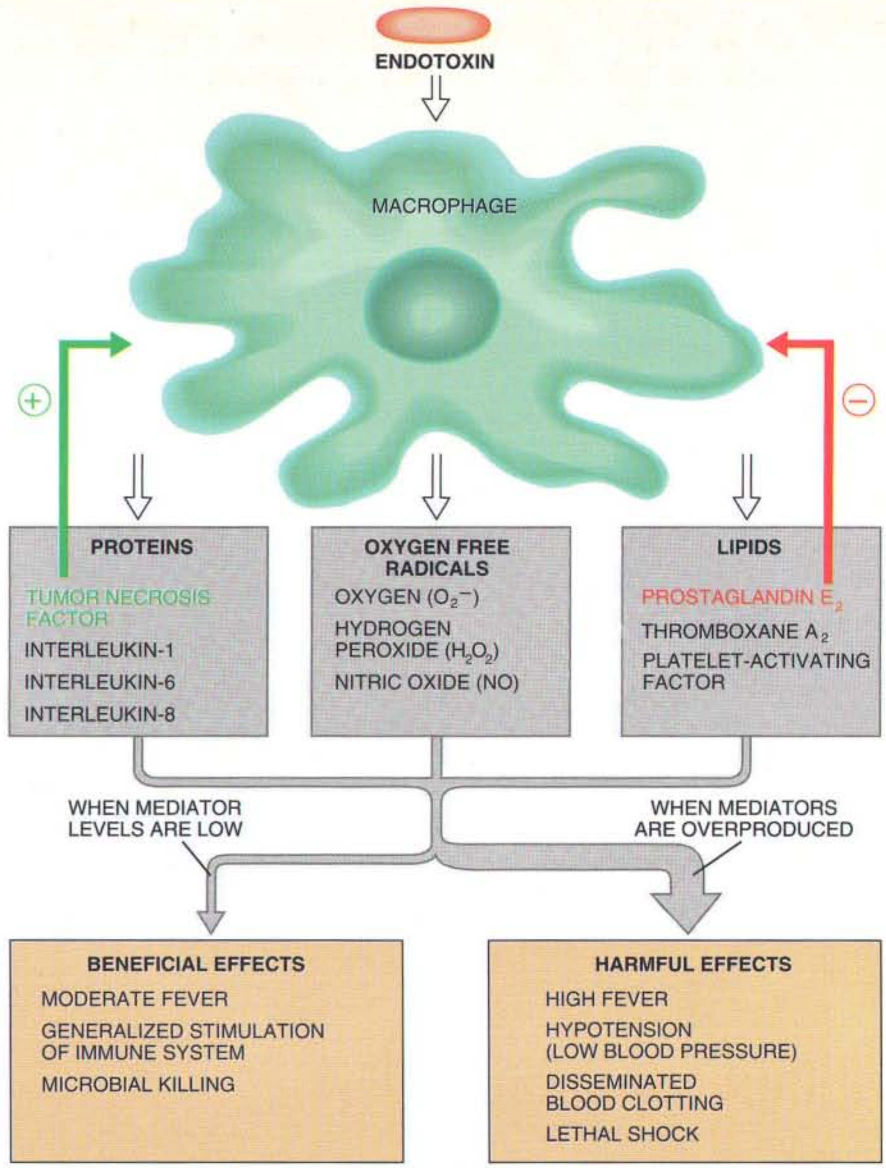
But final proof of the lipid's power emerged only when Tetsuo Shiba and Shoichi Kusumoto and their associates at Osaka University chemically synthesized the full lipid A component of *E. coli* in 1984 and put it at our disposal. Our group, together with that of Lüderitz and Chris Galanos of the Max Planck Institute in Freiburg, then showed that the synthetic version behaved just like natural lipid A. When it was injected into animals, it caused generalized activation of the immune

system, high fever and lethal shock. Three Japanese groups independently obtained similar results.

Successful synthesis of *E. coli* lipid A and demonstration of chemical and biological identity with its natural counterpart finally closed the chapter of research devoted to specifying the endotoxic component of lipopolysaccharide. Yet, because of the structural complexity of lipid A, the possibility remained that some fraction of the lipid, rather than the entire substance, accounted for its activity in the body.

To identify the most critical part, our team and that of Hans-Dieter Flad at our institute studied the influence of a number of synthetic derivatives of *E. coli* lipid A on cultured cells and on animals. For instance, we injected into animals single, instead of the usual paired, glucosamines with their normal fatty acids and phosphates attached. We also tested glucosamine pairs that carried fewer or more fatty acids than normal or a different arrangement of the usual fatty acids.

To our surprise, although some of these compounds were somewhat active, none was more potent than native lipid A. Collectively, these data indicate that, despite its complexity, the entire lipid component of endotoxin molecules, not some particular fraction, is needed for optimal activity. Presumably, the individual elements—the sugars, the phosphates and the fatty acids—combine into a three-dimensional shape that best facilitates interaction with cells of the host.



ENDOTOXINS STIMULATE MACROPHAGES to produce three groups of powerful mediators—proteins, oxygen free radicals and lipids. These mediators may act independently, together or in sequence to engender various effects (*golden boxes*). In addition to its other functions, tumor necrosis factor amplifies mediator synthesis (*green arrow*); prostaglandin E_2 inhibits such synthesis (*red arrow*).

The structural data gathered so far have not yet conclusively revealed the molecular interactions that govern the host's response to lipid A. Nevertheless, research has begun to uncover many details of how endotoxins produce their effects.

It is now known, for instance, that endotoxins must be released from the bacterial surface to be effective. They are set free, as Pfeiffer recognized in 1892, when bacterial cells die, and also, it turns out, when the bacteria multiply. Some bacteria multiply within host cells, others without. Presumably, endotoxins made by the first group must additionally be released from infected cells in order to have an effect.

Once endotoxins are free to act, they do not, as might be expected, kill host cells or evoke other responses directly. They do not, for instance, produce fever by binding to cells in the temperature-regulating center of the brain. Instead, as the Russian-born pathologist Valy Menkin proposed in the late 1940s, when he was working at Temple University, they recruit particular host cells to do their bidding—inducing those cells to secrete mediator molecules. These mediators then act locally or float through the blood, or both, to elicit a diversity of responses.

Studies done by Stephan E. Mergenhagen and his colleagues at the National Institute of Dental Research and, subsequently, by Galanos have since established that macrophages are the recruits most affected. These defensive cells normally take up and destroy any substance that might be harmful to a host. When they become activated, they secrete many different molecules that work in concert, sequentially or independently, to instigate or amplify both specific and nonspecific immune responses against an invader. (Nonspecific components of the immune system, such as macrophages, granulocytes and complement molecules, attack a broad range of invaders.) Some macrophages circulate in the blood; others reside in tissues.

Work in several laboratories has revealed that a small protein called tumor necrosis factor is one of the prime endotoxin mediators made by macrophages. By itself, injected tumor necrosis factor mimics several of the responses attributed to endotoxins, including fever and, if doses are high enough, irreversible shock and death. On the positive side, as its name implies, tumor

necrosis factor can also lure various defensive cells to sites of infection and destroy tumor cells [see "Tumor Necrosis Factor," by Lloyd J. Old; SCIENTIFIC AMERICAN, May 1988].

More recently, endotoxin-stimulated macrophages have been shown to produce the proteins interleukin-1, interleukin-6 and interleukin-8, which exert many of the same effects as tumor necrosis factor. Activated macrophages also release a variety of lipids (some of which contribute to fever and regulate the activity of immune system cells), and they form highly reactive oxygenated compounds known as free radicals. Inside macrophages or on the surface, free radicals contribute to microbial destruction.

Thus, it seems that when a gram-negative bacterium invades tissue and releases moderate amounts of endotoxin there, this array of macrophage products can help eradicate the immediate infection by generating a localized and controlled immune response. The typical effects—mild fever, recruitment of both microbe-specific and less specialized immune components—usually serve recovery and help to protect against other microbial assaults.

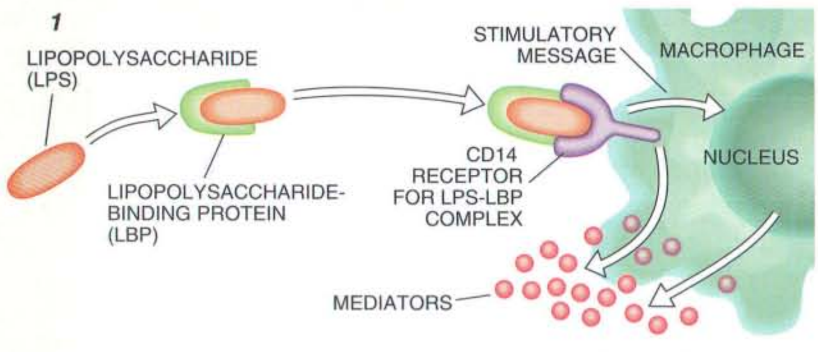
But when an infection is severe and a large amount of endotoxin accumulates in circulating blood, making contact with macrophages throughout the body, systemic release of potent mediators can produce life-threatening shock; as the circulation fails, cells everywhere malfunction and die. Endotoxins can enter the circulation on their own through damaged tissue. Generally, however, lethal effects occur when bacteria themselves gain access to the blood. They multiply rapidly in that medium and, in the process, can liberate huge supplies of toxin to act on macrophages.

Endotoxins, then, are not intrinsically poisonous; their effect depends on the host's response. As Lewis Thomas pointed out in *The Lives of a Cell*, it is the overwhelming, uncontrolled and self-destructive behavior of the host organism that makes endotoxins poisonous. Endotoxins, he wrote, "are read by our tissues as the very worst of bad news. When we sense lipopolysaccharide, we are likely to turn on every defense at our disposal; we will bomb, defoliate, blockade, seal off, and destroy all tissues in the area.... All of this seems unnecessary, panic-driven."

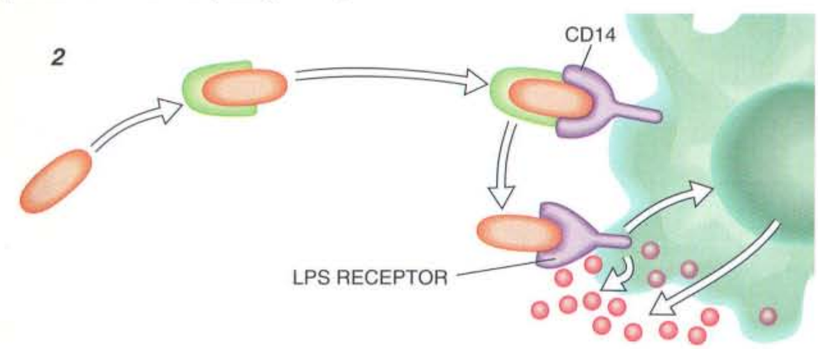
To understand how endotoxins engender such panic, one must learn exactly how they prod macrophages into action. Growing evidence indicates that in the blood endotoxins bind to a circulating molecule known as lipopolysaccharide-binding protein (LBP). Samuel D. Wright of the Rockefeller University and the team of Richard J. Ulevitch and Peter S. Tobias of the Scripps Clinic and Research Foundation in La Jolla, Calif., have shown that the lipopolysaccharide-LBP complex docks with a receptor known as CD14 on the surface of macrophages. This finding makes endotoxins one of the few known substances that interact with a receptor after first forming a complex with a circulating protein.

Less clear is whether the binding of the complex to CD14 alone is sufficient to induce macrophages to produce mediators. It may be that binding to CD14 somehow enables endotoxins to interact with a different receptor, which alone or in collaboration with the CD14 stimulates macrophages. There is some evidence for the latter suggestion. For example, David C. Morrison of the University of Kansas and others have identified molecules on macrophages that

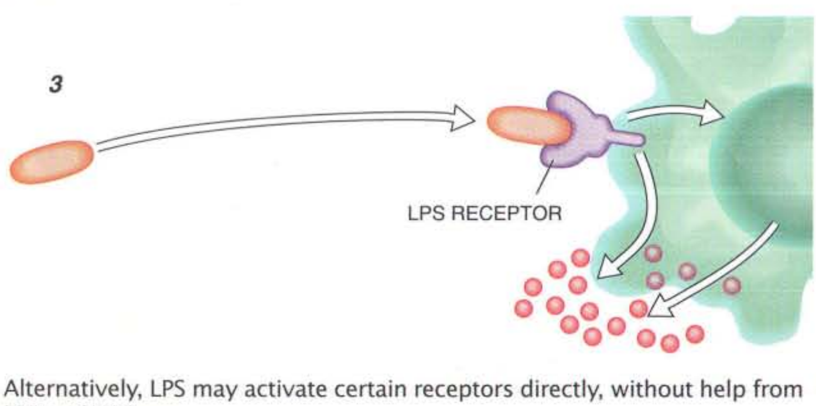
Three Ways Endotoxins May Spur Macrophages to Produce Mediators



Lipopolysaccharide (LPS) (*orange*), another name for endotoxin, binds to the lipopolysaccharide-binding protein (LBP) (*green*) in the blood. Then the complex activates a receptor known as CD14, which instructs macrophages to produce mediators (*red spheres*).



It is possible, though, that CD14 issues no signals but instead enables LPS to activate a second receptor.



Alternatively, LPS may activate certain receptors directly, without help from LBP or CD14.

directly bind endotoxins, particularly the lipid A component. These receptors may take up endotoxins after they attach to the lipopolysaccharide-binding protein and CD14, although it is possible that endotoxins independently activate certain receptors on macrophages in the blood or in tissue.

As the interactions between endotoxins and macrophages come into focus, ideas for impeding those encounters during severe gram-negative infection should emerge. Already investigators are pursuing the discovery that fragments of lipid A, which are less toxic than intact forms (or are even nontoxic), can impede the ability of full lipid A to generate mediator production. Presumably, fragments bind to macrophage receptors and thus deny access to intact endotoxins. The efficacy of partial structures is now being evaluated in animals.

Partial structures or chemically modified versions of lipid A might also prove valuable for enhancing overall immunity in people whose native defenses against infections or tumors are inadequate. Perhaps selected structures will give rise to unusual mixtures of mediators, so that only desirable effects are achieved, such as enhanced nonspecific immunity or destruction of tumors without the complication of shock.

The discovery that host-derived mediators account for symptoms suggests still other approaches to therapy. For instance, one might block the activity or synthesis of the mediators themselves or of the receptors through which they influence cells in, say, neurons and blood vessels. F. Ulrich Schade and Peter Zabel of Borstel find that substances that inhibit the production of tumor necrosis factor or that suppress the action of already synthesized tumor necrosis factor can prevent endotoxins from causing fever in human volunteers and from killing animals.

Consideration of why bacteria make endotoxins has led to additional ideas for treating patients. Gram-negative bacteria may synthesize endotoxins solely to discomfort and threaten their hosts, but we doubt that this is the case. Endotoxin-producing bacteria evolved before susceptible higher organisms, which means the molecules probably originated for some reason unrelated to inducing disease. Furthermore, bacteria do not benefit from disabling or killing their hosts.

Most likely, gram-negative microorganisms make endotoxins simply because they need them. The bacteria apparently require endotoxins in order to reproduce. No gram-negative microbe lacking such molecules has ever been found to replicate well in nature or in the test tube. In addition, structural analyses indicate that gram-negative bacteria rely on endotoxins for protection against external assaults.

Calculations and measurements made by Manfred Kastowsky of the University of Berlin and Harald Labischinski of the Robert Koch Institute in Berlin suggest that the lipid A in *E. coli* loosely resembles a set of wind chimes. The musical "pipes" (the fatty acids), which are embedded in the outermost bacterial membrane, are oriented parallel to one another and perpendicular to the plane of the membrane. The "plate" from which they dangle (the pair of phosphorylated glucosamine sugars) is slightly skewed, tilted at a 45-degree angle relative to the membrane. In this arrangement the fatty acids form hexagonal cables that are packed tightly next to other such cables in the membrane.

Such a rigid, well-ordered structure would explain why gram-negative bacteria are less permeable than the more fluid, phospholipid membranes of human cells and the envelope surrounding gram-positive bacteria. As was first pointed out by Hiroshi Nikaido of the University of California at Berkeley, the lack of permeability would, in turn, explain why many antibiotics that defeat gram-positive bacteria fail to enter gram-negative species or do so only at high doses. Drugs that interfere with the architecture of endotoxins or of endotoxin-containing membranes would therefore help to kill bacteria and make them more vulnerable to antibiotics.

Groups in the U.S. and Sweden are also pursuing ways to inhibit synthesis of lipid A so as to impede production of viable bacterial offspring. They have developed analogues of Kdo for the purpose. Kdo is normally added to a precursor of lipid A before the final fatty acids of the lipid are attached [*see illustration on page 30*]. The analogues inhibit the activity of an enzyme needed for synthesis of this critical sugar. When bacteria encounter these impostors, they fail to make mature lipid A and full lipopolysaccharide. Consequently, they stop growing.

Because Kdo is required in all gram-negative bacteria, the analogues would presumably affect the full range of such bacteria. They would also have another distinction. No existing drugs work by disrupting synthesis of endotoxins; hence, the analogues would represent a novel class of antibiotics.

Therapies may also emerge from efforts to explain why infection by one gram-negative bacterium can boost the power of the immune system to fight other gram-negative species. Recall that Kdo-bound glucosamine pairs occur in all such microorganisms. These conserved regions would be expected to stimulate production of antibodies that recognize the same segment of every bacterium in the class. Drugs composed of such antibodies could have a couple of important benefits.

They would complement the specific immune response triggered by the O-specific chain. Moreover, resistance of gram-negative bacteria to existing antibiotics seems to be increasing, as mutant strains that can evade the drugs gain prominence. Antibodies raised against an essential, conserved region, which cannot change without endangering bacterial survival, might well overcome the resistance problem. Laboratories in many nations are now attempting to develop such a broad-spectrum compound.

Any progress made so far has occurred only because of the combined efforts of investigators in many disciplines, including chemistry, physics, microbiology, genetics, cellular and molecular biology, immunology, pathology, pharmacology and clinical medicine. Similar cooperation will be needed to elucidate further the function of endotoxins in bacteria, to explain how the toxins activate macrophages and induce production of mediators, and to develop new medicines. To promote such work, in 1987 researchers from around the world founded the International Endotoxin Society. This formal unification of forces justifies hope that researchers will soon unlock the remaining secrets of bacterial endotoxins.

FURTHER READING

CELLULAR AND MOLECULAR ASPECTS OF ENDOTOXIN REACTIONS. Alois Nowotny, John J. Spitzer and Elizabeth J. Ziegler in *Endotoxin Research Series*, Vol. 1. Edited by Alois Nowotny. Elsevier Science Publishers, 1990.

GRAM-NEGATIVE ENDOTOXIN: AN EXTRAORDINARY LIPID WITH PROFOUND EFFECTS ON EUKARYOTIC SIGNAL TRANSDUCTION. C.R.H. Raetz, R. J. Ulevitch, S. D. Wright, C. H. Sibley, A. Ding and C. F. Nathan in *FASEB Journal*, Vol. 5, No. 12, pages 2652–2660; September 1991.

BACTERIAL ENDOTOXIN: MOLECULAR RELATIONSHIPS BETWEEN STRUCTURE AND ACTIVITY. E. Th. Rietschel et al. in *Infectious Disease Clinics of North America*, special issue on *Gram-Negative Septicemia and Septic Shock*, Vol. 5, No. 4, pages 753–779; December 1991.

BACTERIAL ENDOTOXIC LIPOPOLYSACCARIDES: Vol. I: MOLECULAR BIOCHEMISTRY AND CELLULAR BIOLOGY; Vol. II: IMMUNOPHARMACOLOGY AND PATHOPHYSIOLOGY. Edited by David C. Morrison and John L. Ryan. CRC Press, 1992.

How Cosmology Became a Science

The discovery of the cosmic microwave background in the 1960s established the big bang theory and made cosmology into an empirical science

by Stephen G. Brush

Did the universe begin, or has it always existed? Scientists long regarded this question as lying outside their concern, in the metaphysical realm of philosophers and theologians. Not until the middle of this century did physicists and astronomers begin to equip themselves with theories powerful enough and experimental techniques sensitive enough to address the issue.

Two competing cosmologies then emerged. One, popularly called the big bang, assumes that the universe evolved from initial conditions so hot and dense that only radiation and elementary particles could exist; the universe then expanded and cooled, precipitating the stars and galaxies. The opposing model offers a universe that has always existed; the dispersal of matter resulting from the observed expansion of the universe is compensated by the continuous creation of matter.

The big bang theory has prevailed, largely because of the prediction, observation and interpretation of a phenomenon known as the cosmic background radiation. This radiation, widely regarded as the afterglow of the big bang, suffuses the sky in all directions at microwave frequencies. Arno A. Penzias and Robert W. Wilson of Bell Laboratories discovered the cosmic background in 1964-65 while trying to rid their radio antenna of microwave noise. The steady state model of the universe predicted no such radiation and could not plausibly account for it. Thus, for the first time, hypotheses about the origin of the cosmos had faced an empirical test that left a winner and a loser.

Rarely do theories stand or fall on the outcome of a single test. This time, however, opinion shifted almost overnight. Within a few years, most cosmologists had either adopted the big bang theory or ceased publishing in the field. Penzias and Wilson won the Nobel Prize in Physics in 1978 for their achievement. Just this past April, measurements of minuscule variations in the background radiation vindicated another prediction of the big bang theory.

Yet no one could have appreciated the significance of the cosmic microwave background without the legacy of knowledge that many other scientists had been building throughout the century. The history of the discovery yields another kind of insight. By following the story past 1965 to see how the discovery affected the standing of rival cosmological theories, we can test competing ideas about the nature of scientific progress.

Big bang cosmology began to come into focus in the 1930s, after Edwin P. Hubble, the eminent American astronomer, showed that galaxies appear to recede from one another and that the most distant ones recede at the greatest rate. Hubble's finding implies that the universe is expanding. It was also interpreted to imply that the cosmos had once been concentrated in a very small space at a definite time. Alexander A. Friedman, a Russian physicist, and Georges Lemaître, a Belgian priest, each used Albert Einstein's general theory of relativity to describe how such an expanding universe might evolve.

Nuclear physics played a role by providing the tools with which to model the synthesis of the elements from fundamental particles. Those tools served not only George Gamow, champion of the big bang, and his colleagues Ralph A. Alpher and Robert Herman but also Fred Hoyle—then at the University of Cambridge—who favored the rival steady state theory.

Vital to the theoretical work was the

COSMIC AUDITORS Arno A. Penzias (*left*) and Robert W. Wilson (*right*) of Bell Laboratories pose on the microwave horn antenna (*shown on this page*) that first cupped an ear to the big bang.

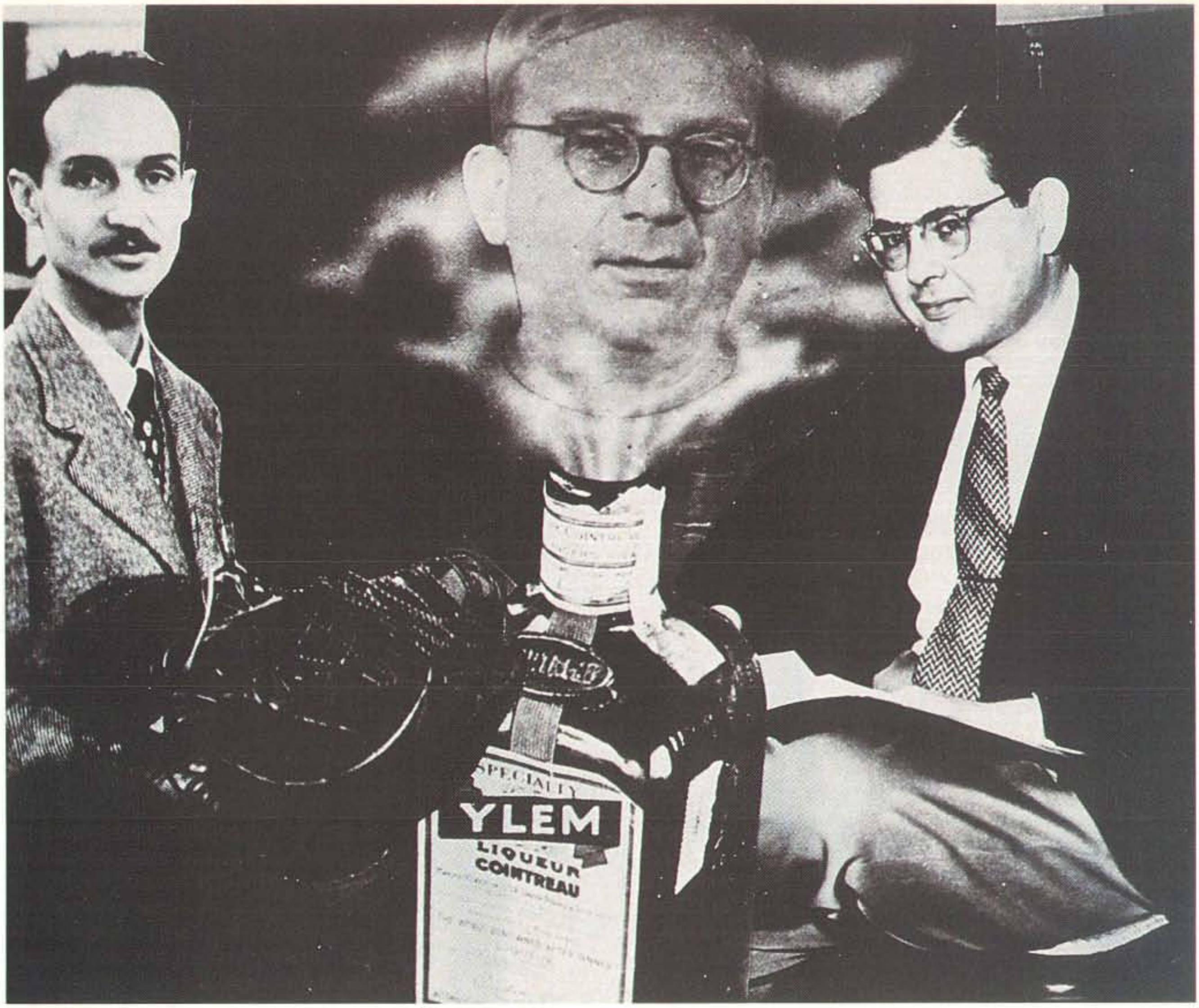
STEPHEN G. BRUSH has taught the history of science at the University of Maryland at College Park since 1968. Before that he conducted research in theoretical physics at Lawrence Livermore Laboratory and helped to write the text for a high school physics course developed at Harvard University. He has written on the development of the kinetic theory of gases, the origin of the solar system and other topics in the history of modern physical science. He is spending the 1992-93 academic year at the Institute for Advanced Study in Princeton, N.J., analyzing how empirical tests affect scientists' choice of theories.

contribution that Einstein and Max Planck made around the turn of the century while formulating the physics of blackbody radiation. The blackbody gets its name from its idealized property of absorbing all incoming radiation and then reradiating it. This reradiated energy is distributed across the spectrum in a highly characteristic pattern, predicted by Planck. Because the primordial fireball, in its early phases, would have put energy and matter into perfect thermal equilibrium, the first radiation liberated from the cooling explosion would have to have displayed the blackbody pattern.

Still to be supplied was a precise calculation of how energetic that spectral pattern would appear today, many billions of years after the fireball began to expand and cool. What was the temperature of the radiation in space? An answer to that question could come only after scientists developed a quantitative theory of the evolution of the fireball after the big bang.

The development of this quantitative theory began with Gamow, a Russian-born physicist who had made his reputation by explaining radioactive decay. In the 1930s he came to the U.S., teaching first at George Washington University and then at the University of Colorado. At George Washington, he concentrated on the astrophysical and cosmological aspects of nuclear reactions—above all, the mechanisms by which the first elements had been synthesized.

Gamow looked for his answer at both ends of the cosmic scale. In the early 1930s astronomers showed that most stars were composed predominantly of hydrogen and helium. It was reasonable to assume that hydrogen was the first element to form because its nucleus contains but a single proton and that helium—the next heaviest element, whose nucleus contains two protons and two neutrons—was the first "higher" element formed by the fusion of hydrogen. But protons will fuse only if some force overcomes the immense electrostatic repulsion between them. This process seemed to require so much heat and pressure that only a primordial event or the interior of a star could have provided the right conditions.



STUFF AND NONSENSE: In a montage he made to amuse friends, George Gamow emerges, genielike, from a bottle of the primeval matter created in the big bang. He is conjured up by Robert Herman (*at left*) and Ralph A. Alpher (*at right*), who showed how such matter—which they called "ylem"—could have combined to form the light elements.

The reigning theory of the nuclear physics of stars, which remains for the most part valid today, had been developed in 1938 by the German-born physicist Hans Bethe of Cornell University. Bethe wanted to explain how the sun shines. He did so by assuming that nuclear fusion in stellar interiors converts mass into energy. Specifically, Bethe proposed that two fusion reactions could take place in stars like the sun: one fuses protons into helium nuclei, and another adds protons to carbon nuclei to form heavier elements.

But where did the carbon originate? That question was not answered until the 1950s, when Hoyle proposed a reaction that could produce carbon from three helium nuclei under the special conditions found at the core of a star. That reaction and others needed to create heavier elements were confirmed experimentally, in a high-energy particle accelerator, by William A. Fowler and his group at the California Institute of Technology. Hoyle and E. E. Salpeter provided important theoretical help. By 1957 a scheme explaining how stars might have synthesized most of the elements from hydrogen and helium had been worked out by Fowler, Hoyle and Margaret Burbidge of Caltech, together with Geoffrey Burbidge, then at the Mount Wilson and Palomar observatories. The work was done independently by A.G.W. Cameron, then at Atomic Energy of Canada. Yet the cosmic abundance of helium remained a mystery.

Gamow had already formulated a daring hypothesis that ultimately led to the solution of the helium puzzle. In his version of the big bang, Gamow suggested that the elements might have formed even before the stars came into being, in a stupendously hot and dense gas of neutrons. Some of the neutrons would then have decayed into protons and electrons—the building blocks of hydrogen. In 1948 Gamow, known for his impatience with detail as well as for his brilliance, assigned the task of developing the theory to Ralph Alpher, a graduate student at George Washington. Alpher later joined forces with Robert Herman of the Johns Hopkins University Applied Physics Laboratory. Alpher gave Gamow's initial substance the name "ylem" from a Greek word meaning "primordial matter."

According to Gamow's theory as worked out by Alpher and Herman, larger nuclei formed in the primeval inferno when smaller ones, beginning with hydrogen, grew through the successive capture of neutrons. The

process continued until the supply of free neutrons ran out, the temperature fell and the particles dispersed.

Hoyle attempted to belittle this new rival to his own steady state scenario by calling it the big bang theory. The attempt at ridicule backfired: the phrase was so vivid that the theory's advocates adopted it as their own.

Alpher and Herman soon realized that the radiation pervading their model universe would maintain the spectrum of a blackbody source as it cooled. Moreover, they could calculate how the expansion of the universe would have attenuated this radiation and reduced its temperature. The two scientists used estimates of the present density of matter in the universe to predict the temperature of the cosmic background radiation today and derived a value of about five kelvins (degrees Celsius above absolute zero).

Astronomers did not rush to confirm the prediction, perhaps because they did not know how to pick out the cosmic background from other radiative sources or perhaps because they did not take seriously the cosmology on which the prediction was based. The original version of the big bang theory had two major drawbacks. First, it failed to explain the formation of the elements beyond helium, which has a mass number of four. Because there are no stable isotopes having mass numbers of five and eight, one cannot make heavier elements out of helium by adding neutrons one at a time. This problem could be solved only by invoking the stellar nucleosynthesis of Hoyle, Fowler and their collaborators, a concept associated with the steady state theory. Indeed, the modern version of the big bang theory assumes that elements beyond helium arose only after the formation of the first generation of stars.

A second objection to a big bang universe involved the question of age. Astronomical measurements of the distances and recessional speeds of galaxies, in conjunction with Hubble's law of expansion, implied that the universe was two billion years old. Yet the rocks of the earth's surface prove that the planet is significantly older than that.

The steady state theory was conceived to resolve this apparent contradiction. One night in 1946, three young scientists in Cambridge, England—Hoyle, Hermann Bondi and Thomas Gold—went to see a ghost story film, *Dead of Night.* As Hoyle later recalled the movie, it "had four separate parts linked ingeniously together in such a way that the film became circular, its end the same as its beginning." Gold asked his friends whether the universe might be similarly constructed. In the ensuing discussion the workers sketched out a dynamic but noncyclic model of the universe that would always look the same even though it is always changing.

According to Hoyle, Bondi and Gold, the universe had no beginning.They argued that the galaxies' rushing away from us does not imply a continuous attenuation of matter: our own galaxy will never be left all alone, they said, because matter is being created continuously, at a rate just sufficient to compensate for the matter that is disappearing from the visible universe. This new matter will eventually form stars and galaxies, so that the universe will always look about the same to any observer at any time.

One might object that the creation of matter out of nothing violates the law of conservation of mass and energy. The riposte is obvious: the big bang also violates this law and does so by creating matter all at once, at the beginning of time, when the act is beyond the reach of scientific study. (In a later version of the steady state theory, Hoyle proposed that gravitational energy creates matter, a refinement that restores the overall conservation of mass-energy but introduces other problems.)

Proponents of the steady state asserted that their theory was more scientific than the big bang because it postulated a process—continuous creation—that might in principle be observed. Moreover, they argued, their theory made definite predictions of a kind that astronomers could test in the near future.

In staking their model on the outcome of a small number of observations, Bondi, Gold and other proponents of the steady state model explicitly invoked the doctrine of Karl Popper, an Austrian-born philosopher now living in England. Popper defines science as a discipline founded on the creation of hypotheses that predict phenomena—preferably new ones—that can be tested. If a prediction fails, the scientist abandons the hypothesis; if the hypothesis survives, the scientist does not claim to have proved it but merely to have established the hypothesis as a basis for further research.

Popper's principle holds that testability rather than truth should be the criterion for judging scientific theories. For instance, Popper dismisses Marxism and psychoanalysis as "pseudoscience" because he believes those theories are so flexible that they can explain any fact and thus elude any test.

Bondi proposed to challenge the steady state theory by comparing the universe as it is with the universe as it once was. Because the steady state theory says the universe always looks the same, it predicts that galaxies formed recently will resemble those formed long ago. If you look out into space—and thus back in time, because the speed of light is finite—and see that distant galaxies are different from nearby ones, Bondi concluded, "then the steady-state theory is stone dead." Like others writing before 1965, however, Bondi failed to mention another test of the steady state model: it does not predict a cosmic microwave background.

The theory failed the test Bondi had set it. In the 1950s and early 1960s a variety of astronomical observations showed that the universe had changed significantly over time. Martin Ryle of Cambridge counted both distant and nearby radio sources, knowing that the more distant signals had taken longer to arrive and thus reflected an earlier stage in cosmic history. Ryle concluded that there had been fewer sources in the past. Although some astronomers argued that he had not proved his case, additional supporting evidence emerged when astronomers discovered what seemed to be the oldest radiative sources—quasistellar objects, or qua-

Two Cosmological Theories and Their Predictions

	TEMPERATURE OF SPACE	AGE OF UNIVERSE	CREATION OF MATTER	FORMATION OF ELEMENTS
BIG BANG	Initially very high; now a few degrees above absolute zero	Two billion years (1950 version); 10 to 20 billion years (current estimate)	All at once	Occurred just after big bang
STEADY STATE	Zero (no radiation)	Infinite	All the time	Occurs in stars all the time

FRED HOYLE, champion of the steady state universe, conceived of the theory with Hermann Bondi and Thomas Gold in 1946, after the three had seen a ghost story film whose plot ended in a return to the opening scene.

of matter. A new shock was needed.

The discovery of the cosmic microwave background provided that shock. Penzias and Wilson made the discovery by measuring the temperature of space or, as a physicist would say, by detecting the Planck blackbody spectral distribution that corresponds to a particular temperature. Electromagnetic radiation pervades the regions between the planets and the stars, and it can be detected by instruments on the earth. Much of this radiation comes in specific frequencies determined by the physical and chemical properties of astronomical sources. It thus cannot be accurately characterized by a single temperature. Instead investigators look for radiation that is in thermal equilibrium at a particular temperature. That is to say, the radiation is continuously distributed over different frequencies according to the law discovered by Planck in 1900.

The Planckian distribution has a characteristic shape for every temperature [*see illustration on opposite page*]. For the universe in which we live, the background radiation corresponds to a temperature slightly less than three kelvins. The distribution peaks at a wavelength of about 0.18 centimeter, which is in the microwave region of the spectrum.

sars. These objects had no contemporary parallel whatsoever.

Meanwhile the awkward issue of the disparity between the age of the universe and the age of the earth was resolved in a way that favored the big bang. In 1952, following the lead of Walter Baade of the Mount Wilson Observatory, astronomers revised their scale of galactic distances upward by a factor of two. The estimated age of the universe therefore doubled. Later work raised it to a minimum of 10 billion years, whereas the age of the earth remained fixed at 4.5 billion years.

Yet many scientists, particularly in Britain, liked the simplicity of the steady state theory and so continued to cling to the concept. They pointed out that one did not have to make arbitrary assumptions about a big bang or worry about what happened before the big bang. Advocates of the steady state model also took heart from the failure of earlier attempted refutations, a record that made them suspicious of any new attacks.

As the steady staters spent ever more time explaining away the evidence accumulating against their theory, their adherence to Popper's methodology steadily became less credible. Instead they seemed to be illustrating Planck's more cynical view of science. Writing in his *Scientific Autobiography and Other Papers* (1949), the great physicist argued, "A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it."

Planck's principle, as historians of science now call it, contradicts Popper's principle by emphasizing the human element in science to the detriment of abstract logic. Just as astronomers can weigh the big bang against the steady state as a description of the universe, so may historians of science try to decide between Planck's and Popper's descriptions of science. Let us see which seems more accurate in this particular case, without undertaking to judge whether science always works in this way.

In 1959 a survey showed that a majority of astronomers rejected continuous creation, although only a third of those voting actually favored the big bang. Even Hoyle abandoned his original model and replaced it with a more complicated hypothesis. In 1964 he concluded that the high abundance of helium in the universe implied it had been "cooked" at temperatures exceeding 10^{10} kelvins. Yet Hoyle refused to abandon the idea of the continuous creation

One can infer a temperature of space indirectly. As Arthur Stanley Eddington pointed out in 1926, the amount of light coming from all the stars—that is, the total energy density—would be equivalent to 3.2 kelvins if converted to thermal equilibrium. But Eddington did not propose a specific procedure for testing his prediction.

At that time, even a scientist of Eddington's caliber would have found the task daunting. Obviously, ordinary thermometers would be swamped by energy coming from the sun, other celestial objects and the earth's atmosphere. Only exceedingly sensitive instruments, tuned to wavelengths between a millimeter and a centimeter and insulated from local sources, can hope to detect the cosmic microwaves.

About 15 years after Eddington made his prescient prediction, Andrew McKellar of the Dominion Astrophysical Observatory in Canada suggested a practical way to measure what he called the effective temperature of space. McKellar, one of the first astronomers to propose that molecules as well as atoms could exist in interstellar space, suggested that the cyanogen (CN) molecule be employed as a thermometer. He noted that cyanogen emits spectral lines whose relative intensity corresponds to the number of electrons in higher-energy states—itself a function of the tem-

perature of space; McKellar estimated that temperature to be 2.3 kelvins.

These indirect approaches could not rule out interference from local sources. To do that, one must detect the radiation itself and map it across the sky. Radar equipment developed at the Massachusetts Institute of Technology during World War II was just barely capable of sensing the cosmic background directly—for anyone who wanted to look for it.

In 1946 a group at M.I.T., led by Robert H. Dicke, reported atmospheric radiation measurements taken by a microwave radiometer. The team noted that the "radiation from cosmic matter at the radiometer wavelengths" was quite sparse—less than the equivalent of 20 kelvins—but did not follow up this observation. Dicke, who subsequently moved to Princeton University, later recalled that "at the time of this measurement we were not thinking of the 'big bang' radiation but only of a possible glow emitted by the most distant galaxies in the universe."

Steven Weinberg, in his book *The First Three Minutes,* suggests two reasons why no one made a systematic search for the background radiation before 1965. First, the big bang had lost some credibility when it failed to explain the formation of elements heavier than helium, so that it did not seem important to test the theory's other predictions. In contrast, nucleosynthesis in stars—a theory linked to steady state cosmology—seemed to explain how heavy elements could have been made from hydrogen and helium, even though it did not explain how the helium had formed in the first place.

Second, Weinberg points to a breakdown of communication between theorists and experimentalists. The theorists did not realize one could observe the radiation with existing equipment, and the experimentalists did not realize the significance of their observations. From this perspective it is noteworthy that Dicke, who is both a theorist and an experimentalist, played a major role: together with P. James E. Peebles, he helped to relate a peculiar microwave noise to cosmological theory.

The most remarkable missed opportunity resulted from a misunderstanding between Gamow and Hoyle. Although each criticized the other's theory, they could still have friendly discussions. In the summer of 1956 Gamow told Hoyle that the universe must be filled with microwave radiation at a temperature of about 50 kelvins. (He arrived at this estimate on his own, after Alpher and Herman had published their prediction.)

As it happened, Hoyle was familiar with McKellar's proposal that the temperature of space is about three kelvins. So Hoyle argued that the temperature could not be as high as Gamow claimed. But neither of them realized that if a direct measurement could confirm the three-kelvin value and also establish the Planckian spectrum, it would refute the steady state theory, which—as Hoyle recognized—predicts a zero temperature for space.

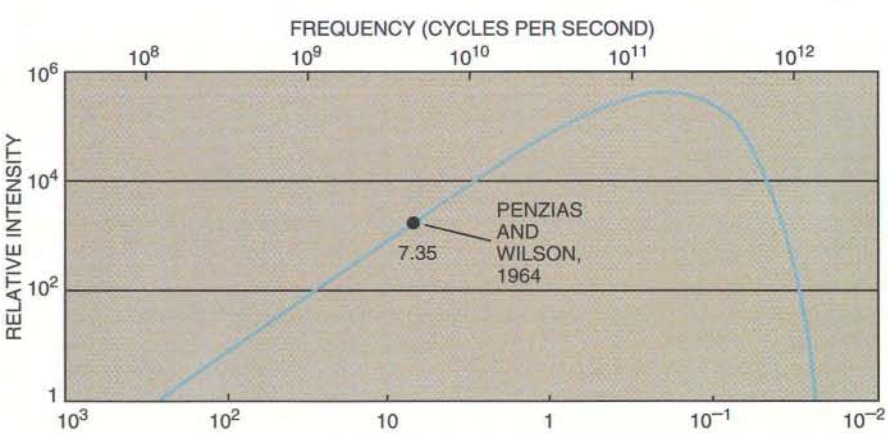
A different kind of communication problem—satellite relays—did lead to the discovery of the cosmic microwave background. Bell Labs wanted its satellites to convey as much information as possible at microwave frequencies, a task that required its workers to find and eliminate noise from all sources. The relay hardware, deriving from the firm's war-related work on radar, consisted of a horn-shaped receiver that Bell Labs engineers Harald T. Friis and A. C. Beck had built in 1942. Another Bell Labs engineer, Arthur B. Crawford, carried the idea much further. In 1960 he built a 20-foot horn receiver at the Crawford Hill facility near Holmdel, N.J. That reflector, originally used to receive signals bounced from a plastic balloon high in the atmosphere, became available for other purposes just in time for Penzias and Wilson.

The two investigators wanted to start a research program in radio astronomy. To prepare the highly sensitive instrument for their work, Penzias and Wilson first had to rid it of microwave noise. They failed in their first few attempts. Finally, in January 1965, Penzias heard that Peebles had a theory that might explain the origin of the stubbornly persistent signal.

Peebles was working with Dicke at Princeton, about 25 miles from the Holmdel laboratory. Dicke rejected the assumption that the universe necessarily began with the big bang. He thought it more likely that the universe went through phases of expansion and contraction. At the end of each contraction, he conjectured, all matter would pass through temperatures and densities intense enough to break down the heavier nuclei into protons and neutrons.

Thus, although Dicke's universe did not start with a big bang, each of its cycles must begin in a similar cataclysm. Moreover, Dicke's cosmology implied an initial fireball of high-temperature radiation that retains its Planckian blackbody character as it cools down, and he estimated that the present temperature of the radiation would be 45 kelvins. He evidently had forgotten his own 1946 measurement that suggested the existence of background radiation at a temperature less than 20 kelvins. Peebles made further calculations from Dicke's theory and obtained an estimate of about 10 kelvins.

Dicke and Peebles, together with two graduate students, P. G. Roll and D. T. Wilkinson, then started to construct an antenna at Princeton to measure the cosmic background radiation. Before they had a chance to get any results, Dicke received a call from Penzias, suggesting they get together to discuss the noise in the Crawford antenna, corresponding to a temperature of about 3.5 kelvins. It was soon apparent that Penzias and Wilson had already detected the radiation predicted by Dicke and Peebles and earlier by Alpher and Herman. But until the two astronomers talked to Dicke and Peebles, they did not know what they had found. The



BLACKBODY SPECTRUM predicted by the big bang theory implies that the earliest radiation in the universe will now appear to be emanating from a source a few degrees above absolute zero. Penzias and Wilson confirmed the spectrum at a single point; others have since confirmed it for a wide range of frequencies.

Selection from George Gamow's *Mr Tompkins in Paperback*

"Your years of toil,"
Said Ryle to Hoyle,
"Are wasted years, believe me.
The steady state
Is out of date
Unless my eyes deceive me,

My telescope
Has dashed your hope;
Your tenets are refuted.
Let me be terse:
Our universe
Grows daily more diluted!"

Said Hoyle, "You quote
Lemaître, I note,
And Gamow. Well, forget them!
That errant gang
And their Big Bang—
Why aid them and abet them?

You see, my friend,
It has no end
And there was no beginning,
As Bondi, Gold,
And I will hold
Until our hair is thinning!"

"Not so!" cried Ryle
With rising bile
And straining at the tether;
"*Far galaxies,*
Are, as one sees,
More tightly packed together!"

"You make me boil!"
Exploded Hoyle,
His statement rearranging;
"*New matter's born*
Each night and morn
The picture is unchanging!"

"Come off it, Hoyle!"
I aim to foil
You yet" (The fun commences)
"And in a while,"
Continued Ryle,
"I'll bring you to your senses!"

Verses by Barbara Gamow

theoretical interpretation was essential to turn mere detection into true discovery. That discovery came more than a decade late because the scientific world had simply overlooked the earlier work by Gamow, Alpher and Herman.

The reports of the groups from Bell Labs and Princeton were sent to the *Astrophysical Journal* in May 1965 and appeared together in the July 1 issue. Publication unleashed a flood of articles in both the mass media and the scientific journals. Even Hoyle admitted that the steady state theory, at least in its original form, "will now have to be discarded," although he later tried to hang on to a modified version that could explain the microwave radiation. But Bondi's emphasis on the testability of the steady state theory had come back to haunt its proponents. Any attempt to twist the theory to explain the new discoveries risked being labeled as pseudoscience.

Although the press was quick to conclude that Penzias and Wilson had confirmed the big bang definitively, scientists realized that their results were limited to only a few wavelengths clustered at one end of the Planck curve. Other explanations of the background radiation, such as a combination of radio sources, could explain those data points. It was not until the mid-1970s that enough measurements at different frequencies had been made to convince the skeptics that the background radiation actually follows Planck's law. The spectrum of the CN molecule played an important part here, as astronomers resurrected and built on the earlier work of McKellar.

By the late 1970s nearly all the original supporters of the steady state model had explicitly abandoned it or simply stopped publishing on the subject. A survey of American astronomers conducted at that time by Carol M. Copp of California State University at Fullerton found that a large majority supported the big bang over the steady state.

The rapid demise of the steady state theory after 1965 shows that Popper's principle, rather than Planck's, applies in this case. The discovery of the cosmic microwave background, combined with arguments about helium abundance and observations of distant radio sources and quasars, convinced most steady staters that their theory was no longer worth pursuing. It had been tried and found wanting.

Yet in 1990, when the steady state theory was all but forgotten, Hoyle and a few of his colleagues tried to revive it as a "mini–big bang" theory, arguing that the evidence does not support the hypothesis that a single explosion created everything. Geoffrey Burbidge recently summarized this view in an essay in these pages [see "Why Only One Big Bang?"; February].

Although proponents of the big bang could dismiss most such criticisms, some puzzles still remained unsolved. For example, the microwave background seemed too smooth. It lacked the slight variations in temperature and, by implication, in density that seemed necessary to seed later gravitational clumping. Without such seeding, there would not have been sufficient time to produce the galaxies and supergalactic structures now observed.

Then, in April of this year, George P. Smoot and his colleagues at the University of California at Berkeley and at the Lawrence Berkeley Laboratory released evidence that may fill this gap in the big bang theory. They announced an analysis of measurements of the cosmic background radiation gathered by an orbiting observatory called the *Cosmic Background Explorer* (*COBE*). The data showed slight temperature variations in the cosmic background, just as had been expected by big bang theorists. The researchers interpret these "ripples" as fluctuations in the density of matter and energy in a very early phase of cosmic history. Such ripples may help explain how matter clumped under the force of its own gravity in time to form the stars, galaxies and larger structures of the contemporary universe.

Did the universe really begin at the big bang, or was there a previous contraction phase—a "big crunch"—that led to the high temperature and density? Will the universe continue to expand forever, or will it eventually collapse into a black hole? Does the creation of the universe involve quantum theory in a fundamental way? These ideas now dominate physical thought [see "Quantum Cosmology and the Creation of the Universe," by Jonathan J. Halliwell; SCIENTIFIC AMERICAN, December 1991]. That scientists consider such questions worthy of serious investigation is itself largely a consequence of the discovery of the cosmic microwave background, which transformed cosmology into an empirical science.

FURTHER READING

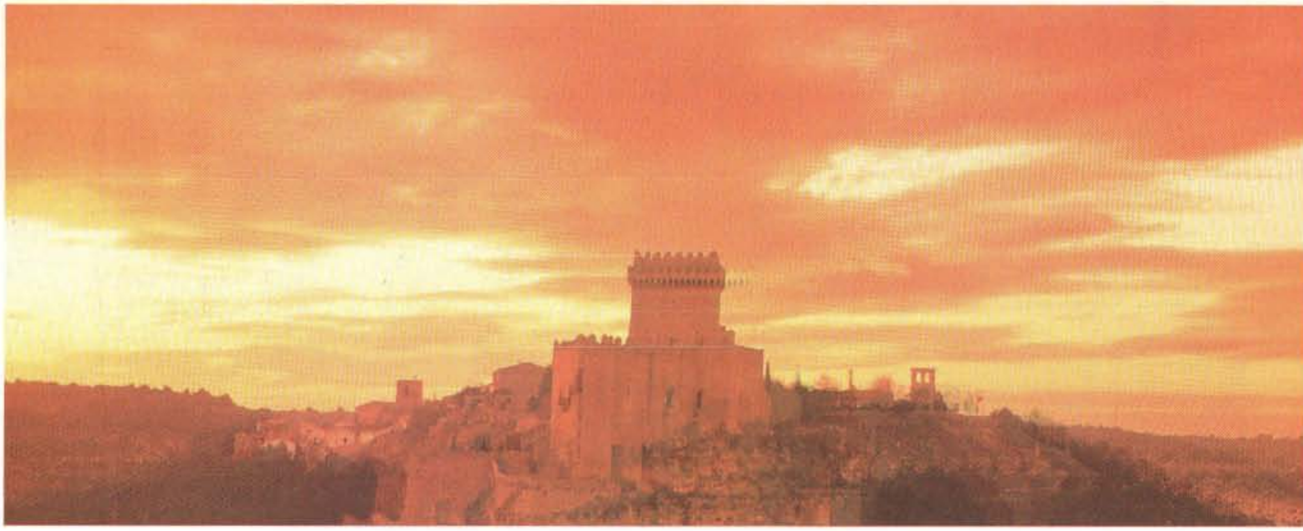
IN SEARCH OF THE BIG BANG: QUANTUM PHYSICS AND COSMOLOGY. John Gribbin. Bantam Books, 1986.

THE COSMIC MICROWAVE BACKGROUND: 25 YEARS LATER. Edited by N. Mandolesi and N. Vittorio. Kluwer Academic Publishers, 1990.

MODERN COSMOLOGY IN RETROSPECT. Edited by B. Bertotti et al. Cambridge University Press, 1990.

THE CASE FOR THE RELATIVISTIC HOT BIG BANG COSMOLOGY. P.J.E. Peebles, D. N. Schramm, E. L. Turner and R. G. Kron in *Nature*, Vol. 352, No. 6338, pages 769–776; August 29, 1991.

"MARQUES DE VILLENA" PARADOR, ALARCON. CUENCA.



We Built them to Keep you Out.
We've Rebuilt them to Keep you In.

Unique

IN 1928, THE DRAWBRIDGE was first lowered on the Spanish *"Parador"*. And for the travelling public, "castles in Spain" became more than a dream. Since then, these state-owned *"inns"* or *"stopping places"* have opened up some of Spain's more remote and most beautiful countryside. Of course, not every Parador is a luxury castle. You could just as easily find yourself staying in a King's palace or a convent. The local leisure opportunities are often as varied as the accommodation. Many Paradors have pools and tennis courts. Some have their own golf courses. While others offer water sports, fishing and even big game hunting right on their doorsteps. Afterwards you can fortify yourself with some of the finest regional cuisine in the land. Some of the recipes are as old as the buildings themselves. Traditionally, Paradors are situated a comfortable day's car-ride apart. And at the end of the day, you can relax still further. Safe in the knowledge that you won't have to fight your way into your hotel.

ESPAÑA

Passion for life.

DELVICO BATES

EXPO'92

1992. The year of the Barcelona **Olympic Games.** The **Universal Exposition** in Sevilla. And Madrid **Cultural Capital** of Europe.

Barcelona'92

Naked Mole Rats

Most of these colonial rodents are nonreproductive workers who attend a fertile queen. In this regard, their social structure closely resembles that of some insects

by Paul W. Sherman, Jennifer U. M. Jarvis and Stanton H. Braude

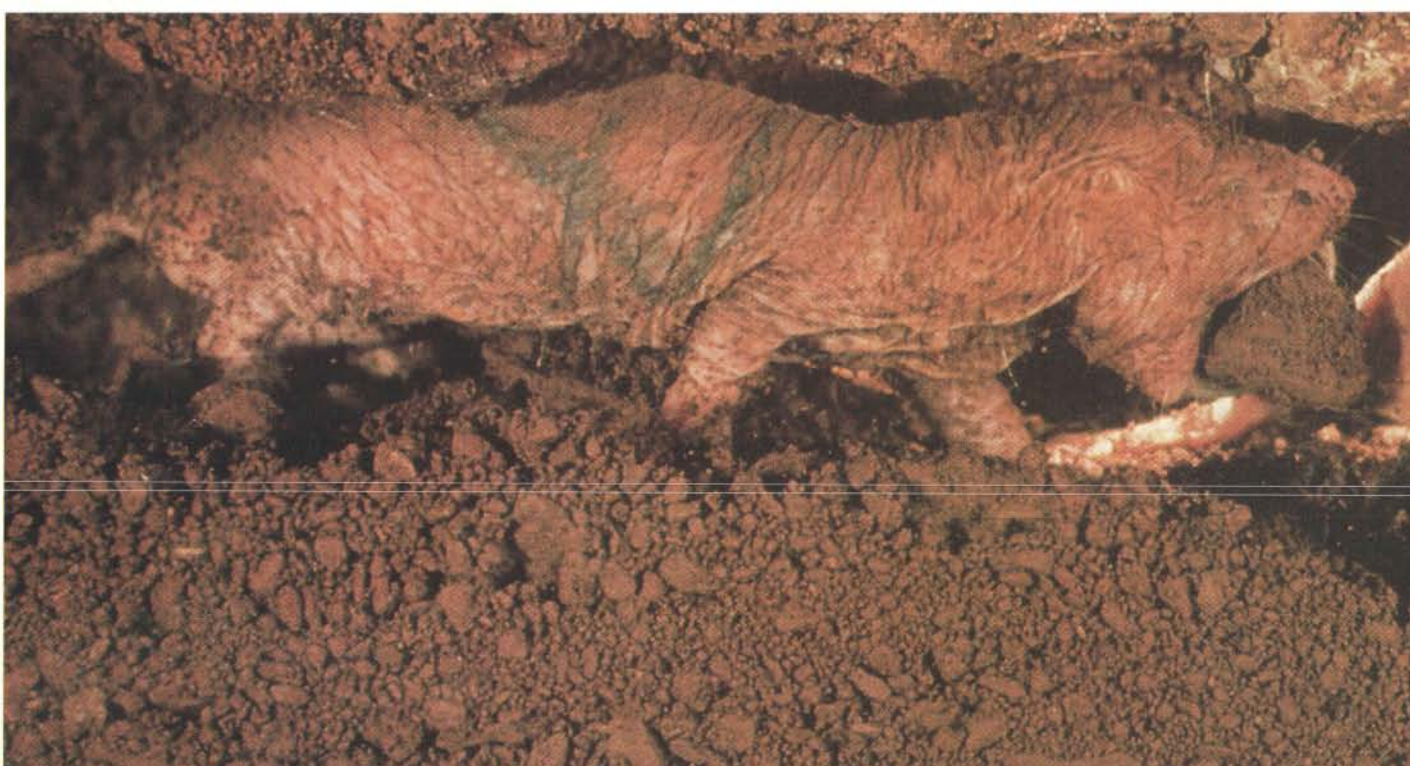
In the rock-hard soil of eastern Africa, networks of subterranean passageways can weave for many kilometers, rivaling Dædalus' labyrinth in complexity. The excavators of these corridors have been described as saber-toothed sausages, as baby walruses and, in more charitable moments, simply as unattractive. We know them as naked mole rats. These hairless, pinkish, buck-toothed rodents are unusual in more than physiognomy. Their social structure resembles that of some insects rather than that of most other mammals. Naked mole rats exhibit eusociality, or "true sociality."

Since the time of Aristotle, people have been fascinated by the altruism, cohesiveness and complexity of the eusocial insect societies—in particular, those of termites and ants as well as certain wasps and bees. Recently eusociality has also been discovered in Japanese aphids and an Australian beetle. Eusocial insect colonies, which may contain several dozen to more than a million members, share three characteristics: at least two generations live together, reproduction is restricted to a few individuals and nonbreeders cooperate to care for the offspring of breeders.

Although cooperative brood care has been observed among many birds and a few mammals, such as wild dogs, eusociality was thought to exist only among the so-called social insects. In the 1980s, however, the discovery that this social system had evolved independently in naked mole rats triggered efforts to understand eusociality in vertebrates and to compare it with eusociality among insects. Studies of naked mole rats offer insights into the ecological and genetic forces that have shaped eusociality, a fascinating evolutionary puzzle.

For Charles Darwin, eusocial insects such as honeybees were potentially the Achilles' heel of his theory of evolution by natural selection. How, he wondered, could nonreproductive castes evolve by gradual steps if they leave no progeny? And how could specialized morphologies and behaviors have been perfected among nonbreeding workers? Darwin proposed a simple and insightful solu-

NAKED MOLE RATS live in subterranean passageways in Kenya, Somalia and Ethiopia. The rodents cooperate to dig their vast network of tunnels as they search for food—the tubers of geophyte plants. One animal gnaws at soil, while others, in turn, transport it to a surface opening to be ejected by a larger colony mate.



tion. He suggested that in this case natural selection acts on entire families as well as on individuals.

In the mid-1960s William Hamilton, a biologist now at the University of Oxford, quantified Darwin's idea by developing a genetic model that became known as kin selection. Hamilton suggested that by helping their parents raise siblings, some of which can breed, nonreproductive workers increase the representation of their own genetic characteristics—in this case, specialized morphologies and helping tendencies.

Because of common ancestry, the siblings that can reproduce carry and pass on alleles for the same traits as their nonreproducing brothers and sisters. By specializing on particular communal tasks, such as caring for young or protecting the colony, nonbreeders increase their parents' reproductive output over what the nonbreeders could produce on their own. Thus, traits beneficial to the survival of the parents and the colony, including worker sterility and specializations for helping, co-evolve even if workers never directly reproduce.

Hamilton's hypothesis helps to clarify a wide array of apparently altruistic behaviors in nature, including eusociality. It also seems to explain the distribution of eusociality in insects. Ants, bees and wasps (order Hymenoptera) have a sex-determining system called haplodiploidy. Males develop from unfertilized eggs and possess only one set of chromosomes: they are haploid.

Females, however, develop from fertilized eggs. Because they receive two sets of chromosomes, one from the egg and one from the sperm, they are diploid. This mechanism of sex determination means that, on average, sisters are three fourths alike: they share all of their father's genes plus half of their mother's. Mothers and daughters are one half alike, whereas brothers and sisters are only one fourth alike. Because a mother has more genes in common with a sister than with her own offspring, female Hymenoptera have a greater stake in rearing sisters.

Textbooks often cite haplodiploidy as the primary explanation for eusociality. But evidence indicates that it is not the whole story. In the first place, the relatedness asymmetries associated with haplodiploidy can occur only when each colony contains a single queen who mates once. But in reality, promiscuous mating and multiple queens exist in many eusocial colonies. These factors cancel the genetic advantages of helping behavior because females are then more closely related to their offspring than to their sisters.

Second, although all species of Hymenoptera are haplodiploid, only a fraction of them are eusocial. To complicate matters further, all termites are eusocial—and diploid. As Malte Andersson of the University of Göteborg in Sweden and Richard D. Alexander and his colleagues at the University of Michigan argue, although kin selection is crucial, haplodiploidy is neither necessary nor sufficient to account for eusociality.

Until recently, eusociality was considered unique to insects. This view changed in 1981, when one of us (Jarvis) reported that naked mole rats live in colonies containing a single reproductive female. The naked mole rat is a diploid rodent with a chromosome number of 60. The mammal was first described in 1842 by a German biologist named Eduard Rüppell as a small, virtually hairless creature from eastern Africa. A glance at the animal and its ample teeth and bulldoglike head proves the aptness of its scientific name: *Heterocephalus glaber,* meaning "different-headed smooth."

PAUL W. SHERMAN, JENNIFER U. M. JARVIS and STANTON H. BRAUDE share an interest in the creatures that have been dubbed saber-toothed sausages. Sherman, who received his Ph.D. from the University of Michigan, is a professor of animal behavior at Cornell University. His research focuses on the social behavior of various mammals and birds. Jarvis received her doctorate from the University of Nairobi and is a professor of zoology at the University of Cape Town. She studies the demography and physiology of African mole rats. Braude, whose Ph.D. is from Michigan, is an adjunct assistant professor at the University of Missouri at St. Louis. He continues longitudinal studies of naked mole rats.

The naked mole rat belongs to the family Bathyergidae. This group of African rodents comprises five genera—three of which are solitary and two social—and about 12 species. The animal's common name is a misnomer. The creature is not entirely hairless, and it is neither a mole (order Insectivora) nor a rat (rodent family Muridae). The evolutionary history of the Bathyergidae is not completely resolved, but it seems the animals are most closely allied to rodents in the suborder Hystricomorpha, which includes guinea pigs, chinchillas and porcupines.

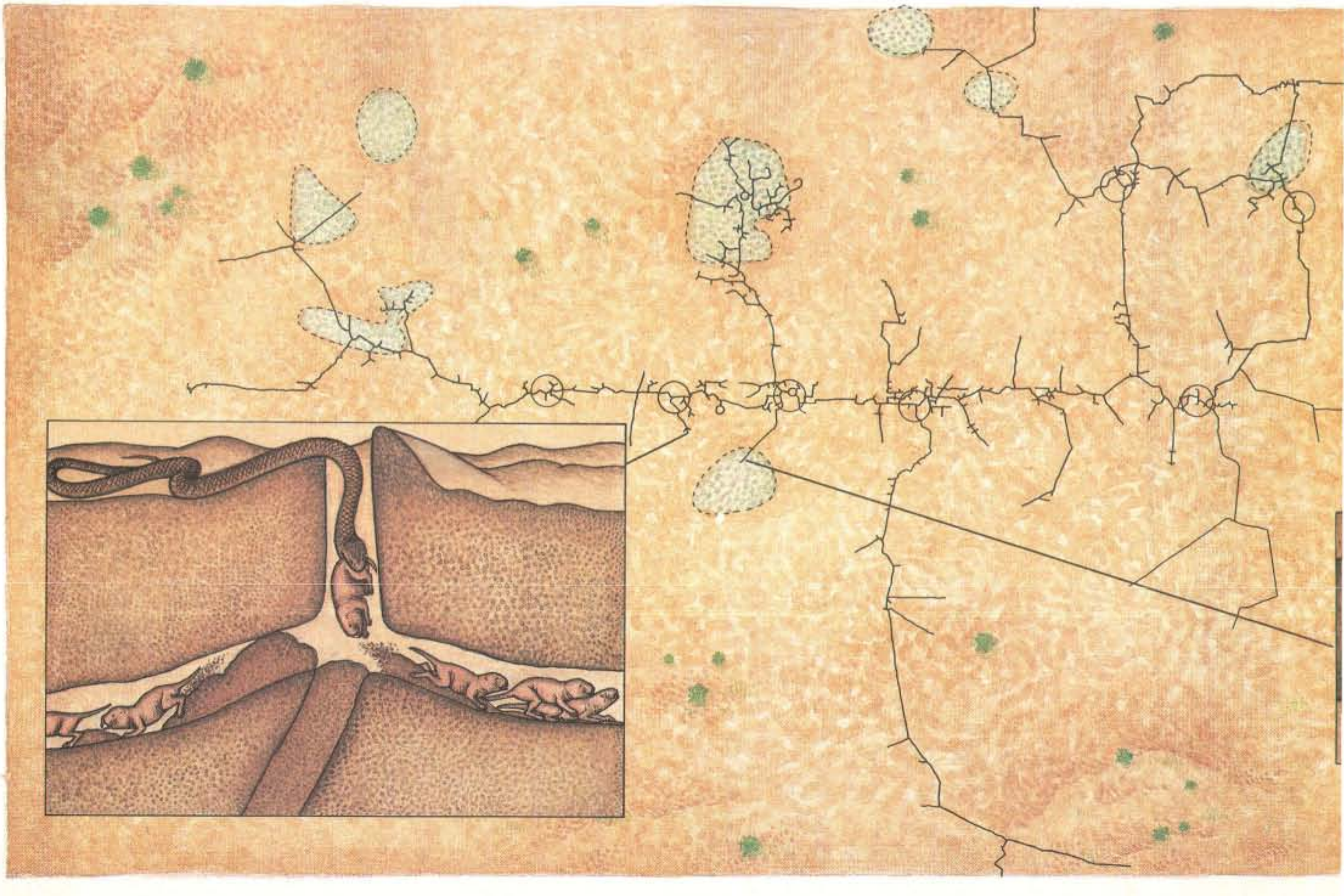
All bathyergids are completely subterranean. They have a variety of specializations for life underground: cylindrical bodies, short appendages, loose skin, minute eyes and tiny ear flaps, as well as huge, protruding incisors for digging. Theodore Grand of the National Zoological Park has discovered that 25 percent of a naked mole rat's muscle mass is concentrated in the jaw region, which allows the rodent to cut through hard terrain, including, in some laboratories, thick plastic and concrete. In contrast, a human jaw contains less than 1 percent of the body's muscle mass; one leg, about 25 percent.

The naked mole rat is the one bathyergid lacking full body fur. Accordingly, the animals are poorly insulated. Individuals, however, do have fringes of hair not only on their lips but also between the toes of their hind feet and on their tail, as well as scattered vibrissae—hairs that serve as sensory organs—on their muzzle. Hairlessness probably co-evolved with group life as a way to minimize cover for ectoparasites, such as mites, which plague all furry social mammals, and as a way to absorb heat efficiently from colony mates.

Unlike any other mammal, a naked mole rat's body temperature fluctuates with ambient temperature. Rochelle Buffenstein and Shlomo Yahav of the University of Witwatersrand in South Africa found that as a result of high heat exchange, small body size and low rates of metabolic heat production, naked mole rats are poikilothermic, or cold-blooded. In nature, the animals live in a relatively thermostable environment: temperatures in the deep tunnels, which are about 50 centimeters underground, remain close to 30 degrees Celsius year-round. Naked mole rats regulate their body temperature behaviorally by basking in warm soil near the surface or by huddling during cold snaps.

The naked mole rat is a prodigious digger. Robert A. Brett, now at the Kenya Wildlife Service, studied a colony of 87 naked mole rats in Kenya's Tsavo West National Park. He recorded that in an average month the colony excavated more than 200 meters of burrows that were four to seven centimeters in diameter. In the process, the animals ejected more than 350 kilograms of soil through some 40 surface openings. All this digging results in intricate systems of connected tunnels and multiple nest chambers that are as big as footballs. The animals move among nests, using the one closest to their current food source.

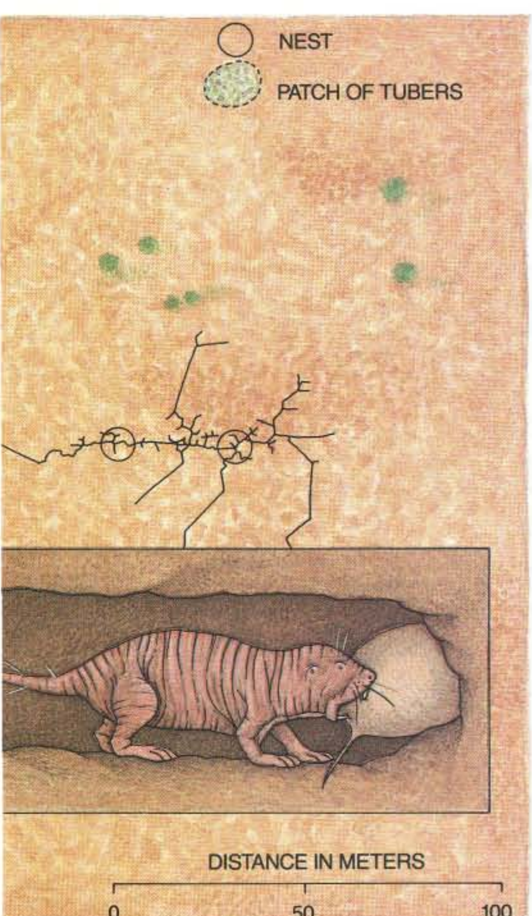
A burrow system can be astonishingly large. Brett put small radio transmitters on a number of individuals in his study colony and tracked them underground. He discovered that the network was more than 3.0 kilometers long and occupied an area greater than 100,000 square meters, about the size of 20 football fields. Colonies carve out these extensive burrow systems in the reddish, iron- and aluminum-rich soils of the semideserts of Kenya, Ethiopia and Somalia. Although they are not rare, naked mole rats are seldom seen. Small volcano-shaped mounds of earth offer



the only clues about their presence. In the cooler parts of the day, these miniature volcanoes "erupt," as the animals kick out soil from recent excavations.

Digging is generally a cooperative endeavor. The animals line up head-to-tail behind an individual who is gnawing on the earth at the end of a developing tunnel. Once a pile of soil has accumulated behind the digger, the next mole rat in line begins transporting it through the tunnel system, often by sweeping it backward with its hind feet. Colony mates stand on tiptoe and allow the earthmover to pass underneath them; then, in turn, they each take their place at the head of the line. When the earthmover finally arrives at a surface opening, it sweeps its load to a large colony mate that has stationed itself there. This "volcanoer" ejects the dirt in a fine spray with powerful kicks of its hind feet, while the smaller worker rejoins the living conveyer belt.

Much of the mole rats' tunneling is done to find food. They eat the succulent tubers of many different geophytes—perennial plants that store water, sugar and starch in swollen roots to enable them to survive the biannual African dry seasons. Most geophytes have irregular or patchy distributions.



TUNNEL NETWORK extends for many kilometers, linking the tubers on which naked mole rats feed (*right inset*). The animals make new nests near their current food source. When mole rats have eaten a large tuber interior, they pack it with soil. Over time, it regenerates. If a snake enters the entrance to a tunnel and attacks a mole rat, workers seal all routes to that area in minutes (*left inset*).

Mole rats forage for them by tunneling blindly, apparently unable to detect the plants through the baked soil. Brett observed that mole rats sometimes eat only the central parts of large tubers and leave the outer layers alone. They then pack the hollowed-out part with soil. In time, the tuber regenerates, and the colony forages from it again.

Where food is abundant, colonies are large and located near one another, about 10 to 20 meters apart, and the mole rats are relatively heavy—averaging 30 to 40 grams apiece. In less productive sites, colony sizes are smaller, the distances between colonies are often more than one kilometer and adult individuals average only 17 to 28 grams.

Like termites, naked mole rats digest cellulose with the help of specialized microorganisms that inhabit their intestine. Also similar to termites, naked mole rats produce two kinds of fecal pellets: one is deposited in a communal toilet chamber; the other is reingested. The latter form is sought and consumed by the breeding female and the young pups. Soft fecal pellets are highly nutritious and laden with microorganisms crucial to digestion.

Naked mole rats are highly colonial. In nature, groups usually contain 75 to 80 animals, but colonies of more than 250 animals have been unearthed. No other bathyergid, indeed no other mammal, consistently lives in such large, integrated social groups within a single burrow system or other domicile. Colonies generally contain more males than females: among 26 colonies, there were 1.4 times more males than females. The reason for this skewed sex ratio is unclear.

Whereas burrow systems and food sources can be studied by excavating them, it is not possible to record the details of naked mole rat social behavior in the field. To facilitate such studies, researchers trap colonies and bring them to laboratories. The animals are housed in artificial tunnel systems maintained in quiet, warm, humid rooms. Each colony member is uniquely tattooed to aid observations of individuals. The mole rats habituate well to captivity, and colonies reproduce regularly; some individuals are still living after 18 years in captivity.

Hundreds of hours of patient observations have revealed that chores are performed by both males and females, but not by all individuals equally. The primary role of breeders is to produce young, to nourish the pups and to keep them clean. Nonbreeders help to clean and carry pups and also to maintain and defend the colony's tunnel system. Labor is divided according to size. Small nonbreeders of both sexes perform maintenance tasks: moving soil, building the colony's nest from bark and rootlets, foraging and transporting food to the nest, and keeping tunnels clear of roots, pebbles and other debris.

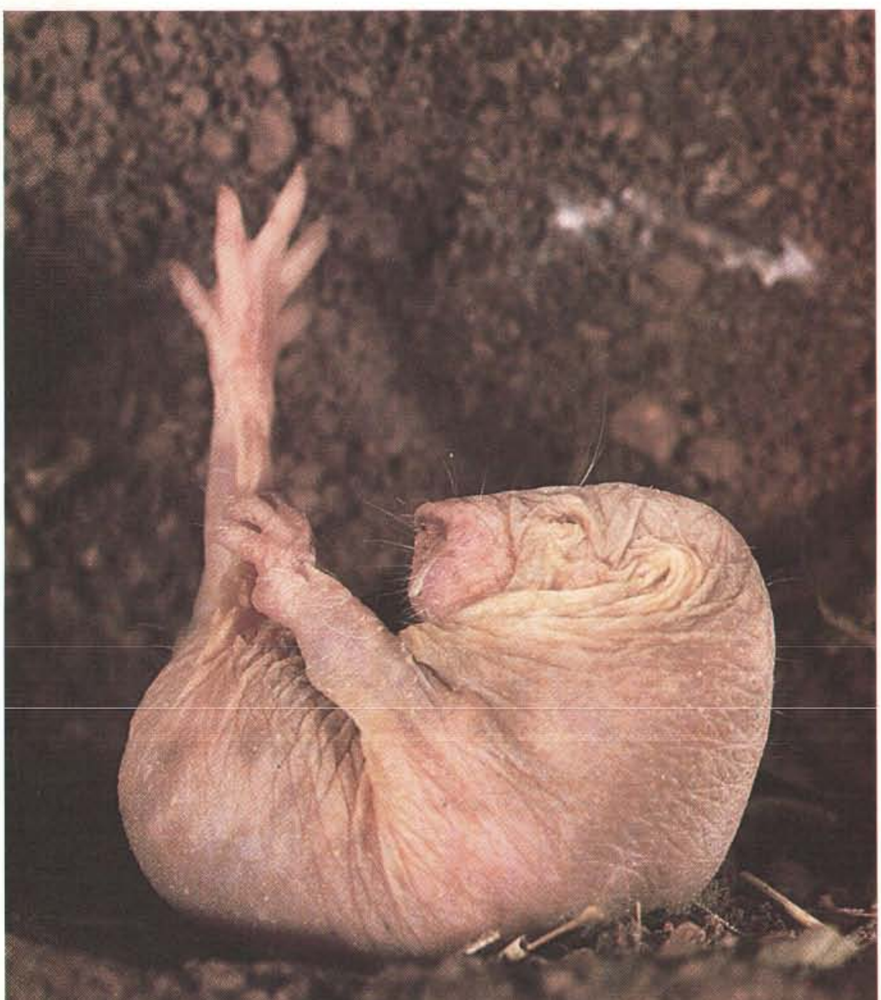
At first, it appeared that large nonbreeders did very little. When they were not excavating new tunnels or ejecting dirt, they spent their time resting in the nest. Then one of us (Sherman) and Eileen A. Lacey, now at the University of California at Davis, discovered that they, too, have a specialized role. If a snake, the mole rat's major predator, enters a volcano, the largest nonbreeders attack and try to kill it or shower it with earth and entomb it. Big nonbreeders also defend their colony against intrusions by members of other colonies.

In some social insects, such as honeybees, behavioral roles change with age, a phenomenon known as polyethism. In naked mole rats, polyethism is based on size. Juveniles that are two to three months old join the colony's work force as maintenance specialists. Later on, the same individuals may become colony defenders and, sometimes, breeders. A mole rat that is an especially rapid and effective forager when it is small and able to squeeze through even the tightest spots becomes a powerful colony defender and volcanoer when larger.

M. Justin O'Riain of the University of Cape Town in South Africa has discovered that some colony members remain small much longer than others. Early social influences and the colony's reproductive history apparently affect a worker's growth rate, its precise role in a colony's work force and its eventual reproductive trajectory.

Communication is obviously necessary for colony integration and coordination, and naked mole rats use several forms of signaling: chemical, tactile and acoustic. John W. Pepper and his colleagues at the University of Michigan and at Cornell University have recorded 17 distinct categories of vocalizations. There are special alarm, recruitment and defense sounds, various contact and aggressive noises as well as calls given only by the breeders—during mating or

NAKED MOLE RAT QUEEN maintains dominance and diligence by pushing or threatening individuals as she patrols the tunnels (*top*). In this instance, the worker held a submissive pose (*bottom*) for several minutes after the royal visit.



urination—or by the pups when they are hungry or distressed. The mole rat's vocal repertoire is the most extensive known among rodents and rivals that of some primates in its richness.

So far we have discussed only the cooperative side of naked mole rat society. Competition, however, lurks just behind the cooperative, communicative facade. Most competitive interactions are subtle, such as mild jostling over preferred sleeping sites, although noisy contests over access to food or digging sites sometimes occur. Michelle Rymond of Michigan and John Schieffelin of Cornell quantified antagonistic interactions and found strong dominance hierarchies within colonies. The queen and the breeding males dominate the nonbreeders, and larger workers dominate smaller ones, regardless of sex.

Sometimes serious conflicts do occur, and the queen is usually the instigator. She is the most active and aggressive individual in the colony, and she frequently patrols her domain, prodding and shoving colony mates. One of us (Sherman) and Hudson K. Reeve, now at Harvard University, discovered that the queen shoves lazy individuals, the large, less related colony members and her mates much more often than she pushes small individuals or close kin. The queen's roughness increases when new resources such as food or nesting material become available, suggesting that her aggression incites colony mates to work. Indeed, shoved individuals do become active and increase their participation in maintenance tasks.

Aggression by the breeding female also maintains her reproductive status. When a queen becomes ill or dies, one or more females suddenly begin gaining weight. Shoving matches occur among the would-be breeders, sometimes escalating into bloody battles, and the females often attack one another when they meet. Calm does not return until one is able to kill, cripple or sufficiently intimidate her rivals.

Conflicts also occur between colonies. When a laboratory colony breaks into another colony's tunnel system, large nonbreeders aggressively defend their home burrows. The defenders confront one another with mouths agape and sometimes clamber onto one another's backs and block the passageway with two tiers of hissing, biting fury. Eventually, the animals attempt to reseal the opening between the burrows with earth. What happens when colonies meet in nature is unknown.

By far the most intriguing aspect of naked mole rat behavior is the re-

striction of breeding. Of 53 captive colonies studied in South Africa, the U.S. and England, 89 percent contained only one breeding female; 93 percent of 26 wild colonies also contained one queen. Clearly, a single queen is the rule, but why two females can sometimes breed at the same time is unknown.

Nonbreeding female mole rats are reproductively suppressed but not permanently sterile. Although most females have underdeveloped reproductive tracts and ovaries, they retain the capacity to breed if the opportunity arises. This ability also characterizes some eusocial insects—particularly those wasps and bees that live in small colonies in which workers have a reasonable numerical probability of reproducing if the queen dies. Physiological sterility characterizes only female ants and termites that live in huge colonies in which the chances of becoming a breeder are minuscule.

In the laboratory, mole rats as young as seven and a half months and as old as eight years have become queens. And in the field, new queens have been recruited from the ranks of both small and large colony members. The switch from worker to queen involves more than just a change in behavior. The queen is usually identifiable as the largest animal in a colony and the only female whose 10 to 14 teats are always prominent. The queen's body is also distinctively elongated. Buffenstein discovered that this shape is caused by the lengthening of individual vertebrae after a female becomes a breeder—a phenomenon unique among mammals. The queen's long body allows her to carry big litters of pups without becoming so wide that she cannot pass through the tunnels.

In contrast to the females, most of the males in a naked mole rat colony have active gonads, although only a few of them mate. When a breeder dies, males between the ages of one and eight years may replace him. Curiously, males seldom fight over the queen when she is in estrus, partly because she shoves fighting males apart and thereby suppresses mating competition. Moreover, a male's chances of breeding in a colony are greater than a female's because the queen will mate with up to three males per estrus period.

Males also begin to deteriorate physically once they start to breed. We do not understand why or how this senescence occurs. But because of the relatively high turnover of breeding males, a male is probably better off waiting for another reproductive opportunity than risking serious injury in a fight over a single mating.

The suppression of breeding is brought about by the queen's behavior—just as it is in several species of bees and wasps. Experiments by Christopher G. Faulkes and his colleagues at the Zoological Society of London have shown that workers will breed if they are isolated from their queen, even if her proximity is suggested by the presence of her soiled bedding and feces. The queen apparently senses subtle changes in the estrogen levels of her workers' urine, a prelude to breeding. Then she selectively shoves would-be breeders, resulting in physiologically detectable stress and anestrus among those subordinates.

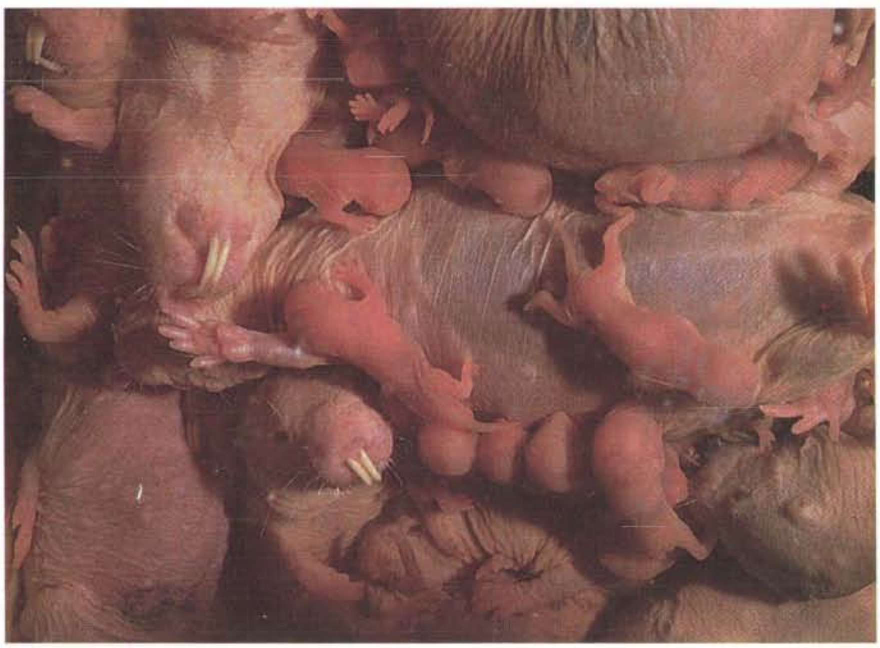
Just before the queen gives birth, when she is bulky and unable to move much, she loses some of her control. As one of us (Jarvis) has shown, the queen is unable to shove females who show hormonal signs of reproductive readiness when she is near parturition. Once the pups are born, the queen rapidly regains control. If she dies during labor, however, her successor will be capable of mating within a few weeks.

After a female becomes a breeder, she bears litters year-round. The queen is sexually receptive four to five times a year. She comes into estrus during lactation, between eight and 11 days after parturition. She then initiates all copulations and mates repeatedly with the breeding male or males.

Gestation lasts about 11 weeks and lactation about four weeks. Just before parturition, the nipples of both male and female colony members often swell. The significance of this event is unclear because nonbreeders never nurse pups. The swelling may indicate the colony's nutritional status or its physiological preparedness for the coming births.

Naked mole rat litters are very large. They average about 14 pups per litter in the field and 12 pups per litter in captivity. One captive queen had 27 pups in a single litter and 108 pups, in five different litters, in a year. The queen can bear so many young because all her needs are met by colony mates.

Once the pups are born, the colony huddles with them in the nest chamber. The pups look like miniature adults and can crawl the day they are born. They are also good climbers, and O'Riain reports they are positively geotactic—meaning they climb upward—enabling them to stay atop the pile of adults. From there they actively seek their mother when she enters the nest. At first, the pups spend most of their time sucking and sleeping. When they are about two weeks old, they begin begging for fecal pellets. The pups first eat geophytes when they are two to three weeks old. By one month of age they are wandering about the burrow system.



NAKED MOLE RAT PUPS are nursed by their mother until they are about a month old. Other colony members typically huddle with the queen and young.

Breeding restriction leads to extreme genetic similarity within a colony. Rodney L. Honeycutt and his colleagues at Texas A&M University determined that within small geographic regions there

MOLE RAT INCISORS allow the creature to gnaw through rock-hard soil. The animal's head region contains 25 percent of its muscle mass. Scattered hairs on its muzzle serve as sensory organs; the rodent has virtually no earflaps.

is little genetic variation within colonies and only slightly more variation between colonies.

Using DNA fingerprinting, Faulkes, Reeve and their colleagues also investigated the genetic structure of colonies. They discovered that colony mates are much more alike than nonkin in other wild vertebrates. For example, relatedness among colony mates averages 0.81, meaning they are considerably closer than full siblings in outbred species, which have an average relatedness of 0.50. Such high relatedness is attributed to descent from the same parents and to juveniles' remaining in their natal colony to breed rather than dispersing. Indeed, an estimated 85 percent of matings in naked mole rats are between parents and offspring or siblings.

Between colonies that are located farther apart, however, there is considerably more genetic divergence. For example, colonies from northern and southern Kenya, separated by about 300 kilometers, differ as much as subspecies of some other rodents. Similar genetic divergence has been reported in several subterranean mammals and apparently results from a combination of a sedentary way of life and extinctions as well as recolonializations from nearby groups. Because of their relative lack of genetic variability, naked mole rat colonies may be particularly susceptible to being wiped out by new diseases.

Since 1986 one of us (Braude) has captured, marked and released about 4,000 individuals from 30 colonies in Kenya's Meru National Park. The animals were captured with the aid of a tunnel-shaped, electronically triggered trap. By recapturing colonies yearly, Braude found that individuals vanish at a surprisingly rapid rate. Indeed, few are recaptured after more than a year. In six colonies studied for up to four years, 21 to 80 percent of the initial residents were present one year later, but only 2 to 15 percent were present after two years, and just 1 to 2 percent remained after three years.

Interestingly, the likelihood that a queen would remain from one year to the next was double that of a worker: 93 percent as opposed to 43 percent. Direct sightings of snake attacks, the discovery of mole rats in the stomachs of snakes and observations of bite marks and missing limbs on some captured animals indicate that predation is an important cause of mortality.

Successful dispersal may be rare. Despite extensive searches, only a few of hundreds of marked animals that disappeared from colonies were relocated in any other colony. It is unclear precisely how new colonies are formed. On the one hand, Brett suggests that large colonies fission to form new ones, based on the relatively large size of even the smallest colony (25 individuals) he unearthed. Colony fissioning has been observed in Sherman's laboratory, and in 1991 one of us (Braude) actually witnessed the fissioning of a colony in nature. On the other hand, discovery of one solitary pair of mole rats far from any possible source colony suggests that the animals may disperse overland.

Evolutionary explanations for insect eusociality have focused on genetic factors such as haplodiploidy. In contrast, explanations for cooperative breeding in mammals and birds have focused on ecological factors. The latter theories emphasize the role of a limited breeding habitat that forces the young to remain at home rather than dispersing and breeding independently. And yet, clearly, naked mole rat society closely parallels that of many social insects.

Our studies help to bridge the theoretical gap between explanations for vertebrate and invertebrate eusociality. We hypothesize that naked mole rats live in groups because of several ecological factors. The harsh environment, patchy food distribution and the difficulty of burrowing when the soil is dry and hard, as well as intense predation, make dispersal and independent breeding almost impossible.

By cooperating to build, maintain and defend a food-rich subterranean fortress, each mole rat enhances its own survival. Once a young naked mole rat stays at home, its personal reproduction is restricted by its powerful mother. Thus, an individual's only reproductive option is rearing siblings. Since the queen specializes in reproduction, many more young are born to a colony than could be born to an individual on its own. Because of inbreeding, siblings are extremely closely related, and thus, as noted by Hamilton for eusocial Hymenoptera, a worker mole rat reaps genetic returns by helping them.

The existence of such extreme altruism and cooperation has led some contemporary biologists to consider eusociality a pinnacle of social evolution. As the eusociality of naked mole rats demonstrates, ugliness is only skin deep.

FURTHER READING

EUSOCIALITY IN A MAMMAL: COOPERATIVE BREEDING IN NAKED MOLE-RAT COLONIES. Jennifer U. M. Jarvis in *Science*, Vol. 212, No. 4494, pages 571–573; May 1, 1981.

THE EVOLUTION OF EUSOCIALITY. Malte Andersson in *Annual Review of Ecology and Systematics*, Vol. 15, pages 165–189; 1984.

THE BIOLOGY OF THE NAKED MOLE-RAT. Edited by Paul W. Sherman, Jennifer U. M. Jarvis and Richard D. Alexander. Princeton University Press, 1991.

Evolution Comes to Life

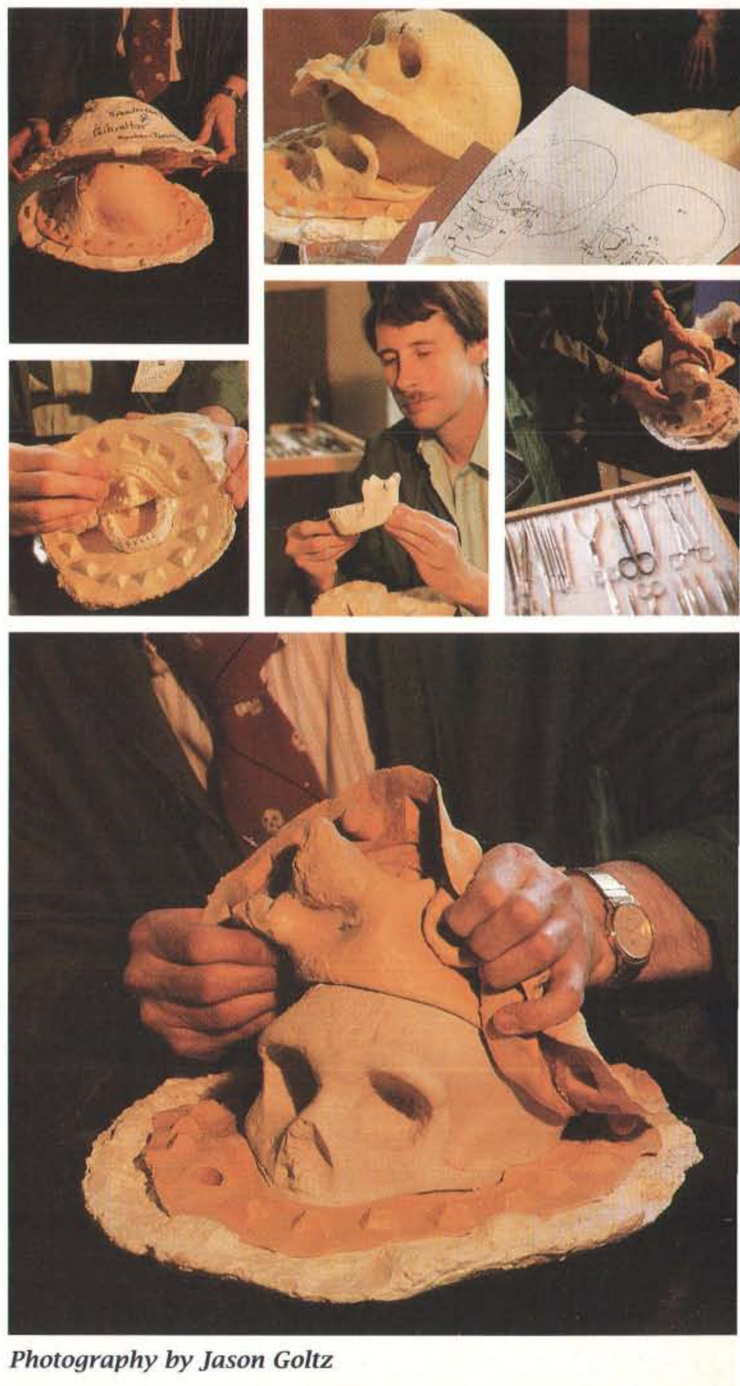
Reconstructing extinct humans as they were in life poses many unfamiliar and awkward problems for paleontologists accustomed to interpreting evolution from bones and teeth alone

by Ian Tattersall

Ancient bones are the objective evidence of biological history. From my standpoint as a paleontologist, they are vastly more informative about extinct creatures than reconstructions or models, in whose creation art plays at least as great a role as science. Yet I am also a museum curator, and from that perspective I am keenly aware that nothing brings the past alive in the public's eye like a well-crafted reconstruction. For the average person, fossil bones are static things: beautiful or majestic, perhaps, but hard to imbue with the attributes of a living, breathing form.

When I was given the responsibility of curating the American Museum of Natural History's new Hall of Human Biology and Evolution, it was therefore evident to me and to Willard Whitson, the designer of the hall, that we needed to include some reconstructions of early humans in the exhibition. Furthermore, we wanted to portray these figures dynamically in the context of situations that our ancestors might have faced long ago. Only thus, we thought, could we truly bring these long-departed relatives back to some semblance of life. We hoped that clever sculpting and modern casting materials could provide us with a level of realism rivaling that of the spectacular dioramas of modern animals in the adjacent galleries.

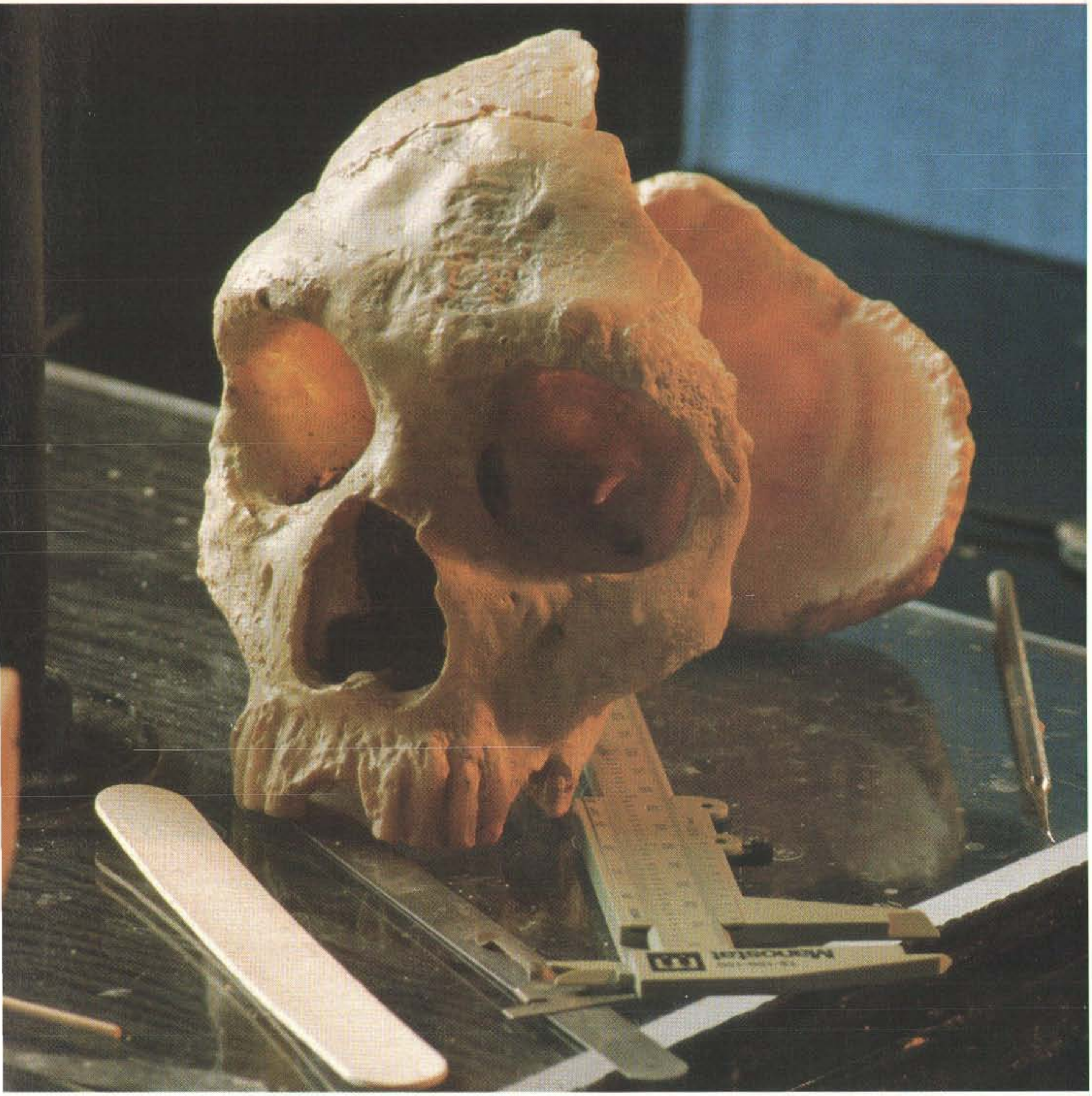
What I had failed to consider, however, was the extraordinary number of awkward decisions that would become necessary as work progressed. We scientists customarily deal with objective matters, and we are happiest when reaching judgments based on testable reality. Untestable speculation makes most of us acutely uncomfortable. As our experience

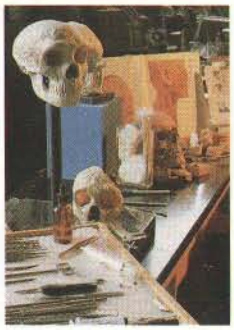


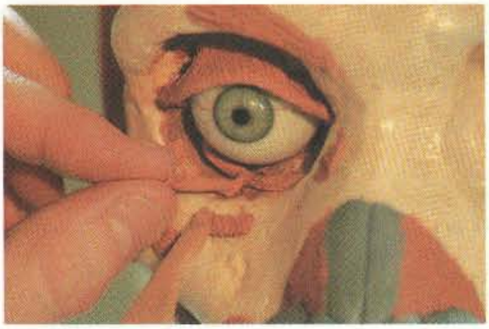
NEANDERTHAL SKULLS and bones offer the only direct physical evidence for the appearance of that extinct people. Working from replicas of fragmented fossils, anthropologist Gary Sawyer sculpts a complete skull and makes a mold of it. A new skull cast from this mold serves as the foundation for further reconstruction work.

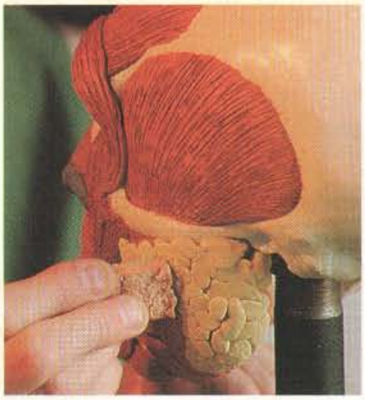
Photography by Jason Goltz

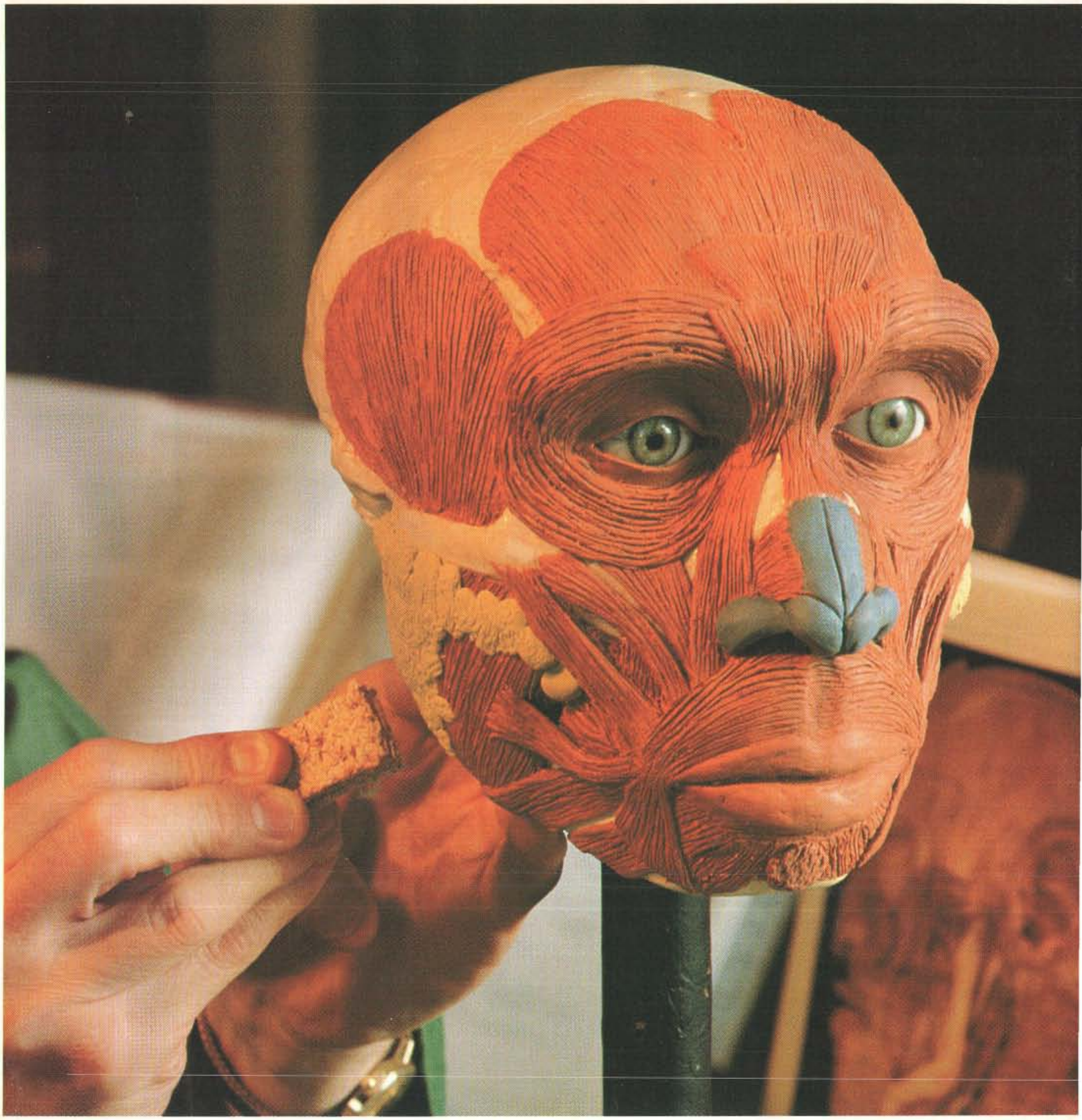
IAN TATTERSALL is a paleontologist and primatologist who has worked extensively on the living and subfossil lemurs of Madagascar as well as on a range of problems in early primate and human evolution. Curator and chairman in the department of anthropology of the American Museum of Natural History, Tattersall is especially interested in the integration of evolutionary theory with the fossil record. He has also served as curator for several major exhibitions at the American Museum, the most recent of these being the Hall of Human Biology and Evolution, scheduled to open in early 1993.







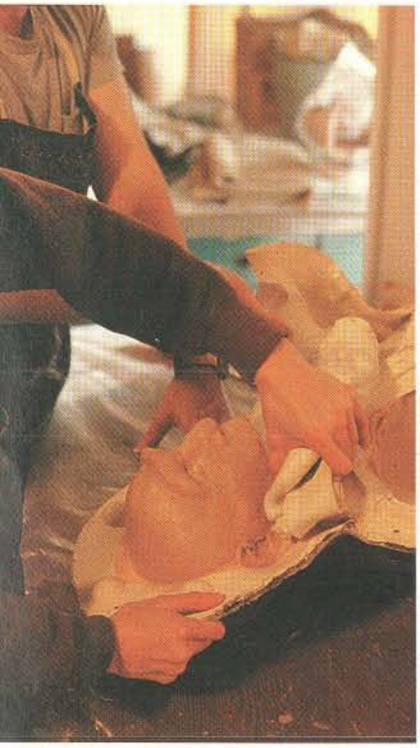








with preparing the exhibits for the hall illustrates, such speculations are inescapable in any attempt to depict extinct human species.

We had decided to re-create three distinct species of human ancestor, reaching back several million years in evolution. The most recent of these is *Homo neanderthalensis,* shown at a site in France where Neanderthals lived about 50,000 years ago. (Although I and many other paleontologists are confident that Neanderthals were a species apart from modern *Homo sapiens,* some workers still dispute the point.) Despite their geologic youth and general similarity to us, the Neanderthals presented a typical set of difficult decisions.

The first choice to be made—what the three individuals in the scene should be doing—was the easiest because archaeologists have learned a fair amount about how Neanderthals lived. For example, the characteristic wear on their stone tools tells us that Neanderthals used flints to cut wood and scrape hides. We settled on showing a male sharpening a wooden spear while a young female scraped a hide and an older female offered advice. Because Neanderthal front teeth are usually very heavily worn, we felt safe showing the young woman with one end of the hide held in her teeth.

For realism in the bodies of our Neanderthals, we chose to cast the bodies of living human models. Neanderthals were more robust than modern humans, but their body proportions were generally similar to ours, and sculptural modifications could accommodate any slight differences between them and us—their longer thumbs and the shape of their shoulder blades, for instance—that might have been detectable in the flesh.

Paleontologists have found a reasonably large number of fairly complete Neanderthal skulls that well convey the general dimensions of the head. Neanderthals had a braincase as big as ours, but it was long, low and lacked the modern high forehead. Their face protruded along its vertical midline: the nasal orifice was large, and the cheekbones receded. Heavy bony ridges overhung the eye sockets.

To determine the appearance of the features overlying the skull, my museum colleagues Gary Sawyer and Steve Brois and I faced some unfamiliar problems. Modeling an extinct human depends on skills much like those of the forensic artists who reconstruct the faces of unidentified murder victims. The key is that muscles leave telltale impressions on the bones to which they attach. From the size and depth of these impressions, one can deduce the dimensions of the muscles. On a cast of the dry skull, an artist or paleontologist can then reconstitute the overlying soft tissues layer by layer, starting with the deep tissues and proceeding outward.

Forensic artists, however, have a great advantage: they can study the soft anatomy of living people. Paleontologists have no living Neanderthal standards for comparison. As a result, the shapes of the nose, lips, ears and other features that are so vital for establishing the character of a Neanderthal face

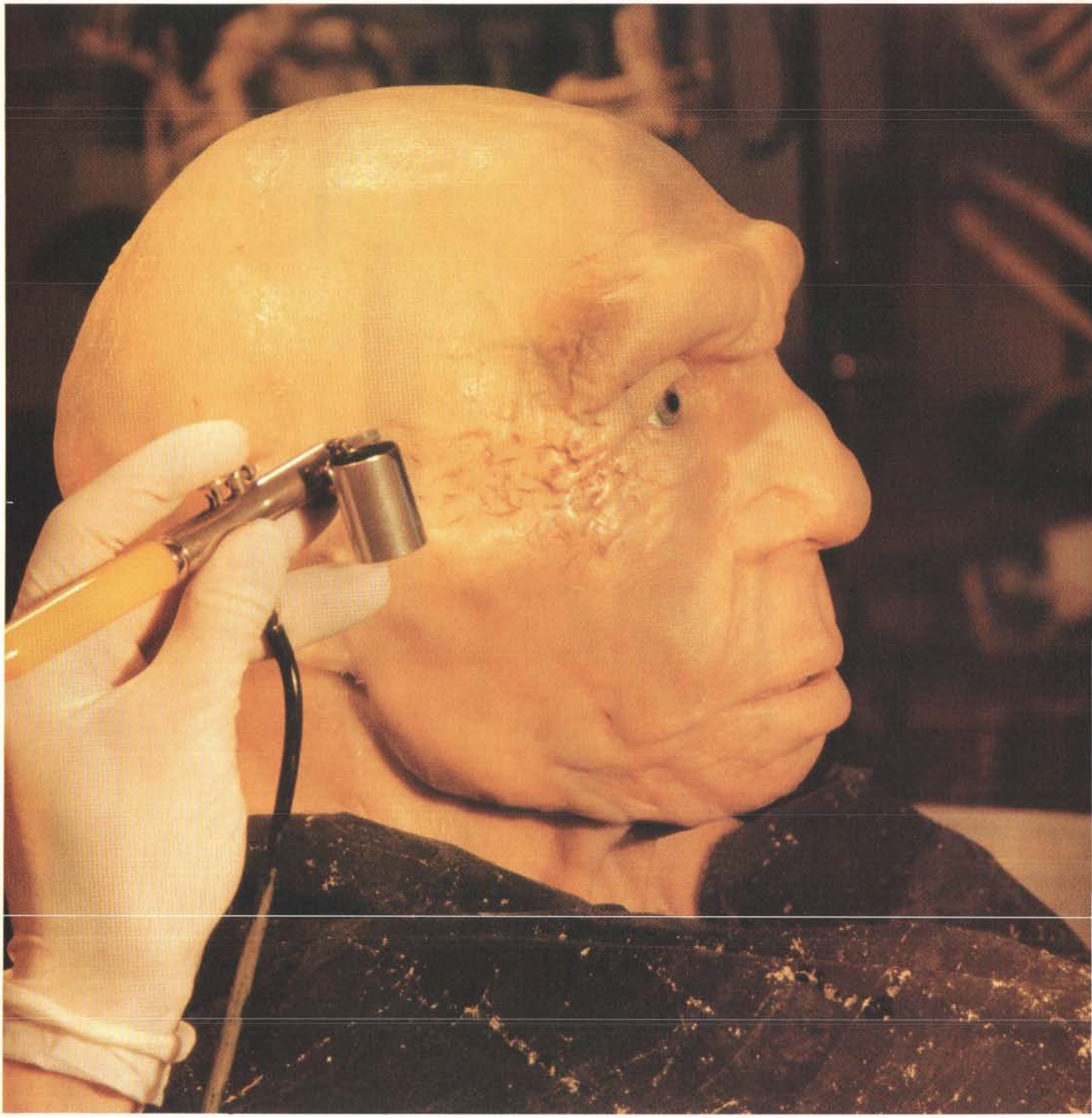
MUSCLES, fat, eyes, ears and other features are carefully added to the replica of the Neanderthal skull, building up from the bone. Although no fossils of these tissues have survived, their proportions can often be inferred from marks on the skull. A plastic coating is eventually applied to create a skinlike surface over the head. Plaster life casts of modern people help to model the body. (The casting process for a hand is shown.) The reconstructed head and the body casts do not become parts of the finished figure. Instead preparator Cathy Leone and other artists use them as the basis for an anatomically accurate mold of a complete Neanderthal. The plastic figure for the exhibit is cast from this mold.

boil down to educated guesswork. In our models, we tried to avoid unreasonably exaggerating these soft tissue features.

Our arbitrary decisions were still only beginning. Once I would have laughed if anyone had predicted that I would spend weeks agonizing over Neanderthal eyebrows. Did they even have eyebrows? (Our closest living nonhuman relatives, the chimpanzees, do not.) If Neanderthals did have eyebrows, where were they on those bulbous browridges? Similarly, how long would untended Neanderthal beards have grown? How much body hair did the men and women have? What was its color and texture? What was the skin color? All these details offered endless scope for quibbling.

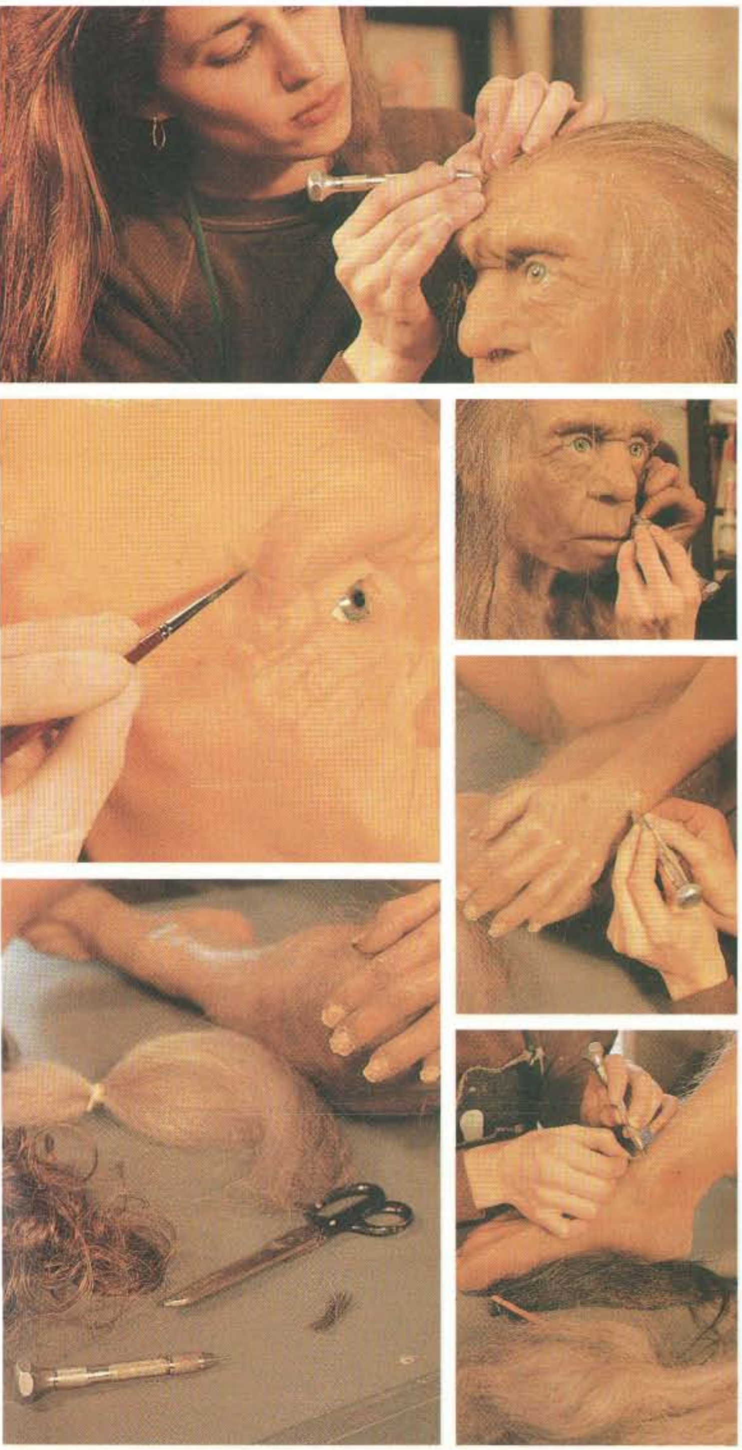
We based our final decisions on the best information available. The high-latitude habitat of these people suggested that Cathy Leone and our other preparators should give them a light skin color like that of modern Europeans. Because the European climate was cold long before the peak of the last Ice Age, about 18,000 years ago, we also felt she should give them body hair well toward the heavy end of the modern range.

Neanderthal humans knew how to prepare hides, and it seems virtually certain that they wore hide clothing, but of what kind? These people became extinct almost 10,000 years before the earliest known bone needles were in use, so neatly tailored clothing was clearly out of the question for them. That exclusion still left many possibilities. Basic is-



sues, such as whether fur was worn on the inside or the outside of the clothing or whether it was scraped off altogether, were unanswerable.

In the absence of scientific solutions, aesthetics was allowed to take over. During the months of work, all of us involved in the project grew accustomed to the evolving appearance of each reconstructed figure. The first major shock always came with the addition of hair to a formerly glabrous head—an old friend was suddenly transformed into a stranger. Draping the naked figures with animal hides that obscured the results of hundreds of hours of laborious body detailing was equally jarring for us. Predictably, perhaps, our Neanderthals ended up with rather minimalist clothing.



Clothing was not an issue in the second diorama, which is set in the Turkana basin of northern Kenya about 1.7 million years ago. That tableau features a close relative of *Homo erectus* that should probably be called *Homo ergaster,* "working man." It was given this name because it used tools, although it was not the first human to do so. Deciding what activities to depict in this scene was more problematic than for the Neanderthals because so much less is known about the behavior of earlier humans.

We had room in the diorama for only two figures, and we chose to use one male and one female. We did not want to suggest that they represented a nuclear family, however, because there is no evidence that these early people lived in such groupings. Compounding our difficulties was archaeologists' uncertainty about whether *H. erectus* and its relatives hunted anything bigger than the smallest game. We wanted to show a carcass being butchered with crude stone implements—we know these people did that much—but we did not want to imply that our protagonists had necessarily made the kill.

An exhibit label might list as many caveats as we liked, but a concrete representation of any behavior would inevitably seem to make a rather definitive statement. In the end, we chose to show the female warding off a jackal with a piece of bone while the male, attacking an impala carcass with a stone flake, is startled by a vulture overhead.

The figures themselves were a little less problematic: from the neck down these early people were quite similar to us. We gave them a dark skin for protection against the tropic sun and sparse body hair, which would have facilitated cooling through the evaporation of sweat. Reconstructing the faces of these small-brained people was trickier than doing so for the Neanderthals because their evolutionary separation from modern humans is so much greater. Nevertheless, the challenges were similar.

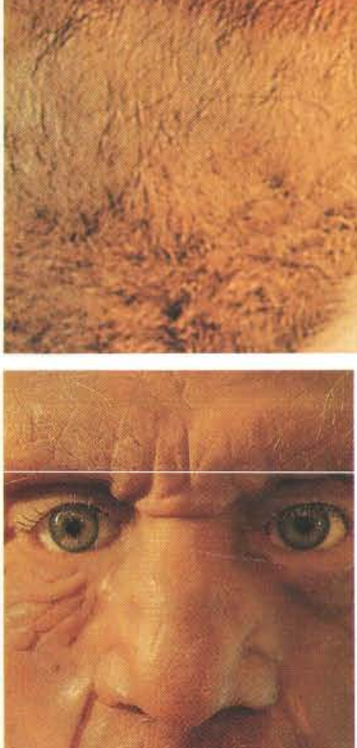
The third diorama takes museum visitors even further back in time. It depicts the making of the famous footprint trails at the Tanzanian site of Laetoli about 3.5 million years ago. A nearby volcano had puffed out a cloud of ash that settled on the landscape and was dampened by rain. Across this muddy surface walked two early humans and possibly a third (although that idea is disputed). The variety of hominoid that made the prints is also disputed, but the only known candidate is *Australopithecus afarensis,* the earliest member of the human lineage yet described.

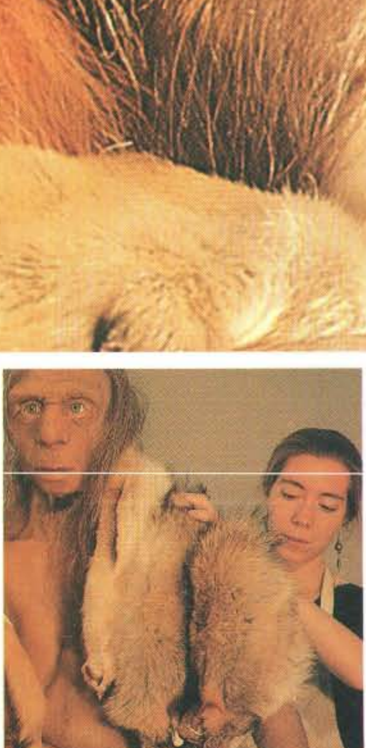
The prints themselves, hardened in the mud and miraculously preserved, show an astonishingly modern upright bipedal gait. The two parallel tracks are those of a large and a small individual walking in step and so close to one another that they must have been in physical contact. Were they an adult and an adolescent, or perhaps a male and a female? Fossils show that *A. afarensis* males were much larger than the females.

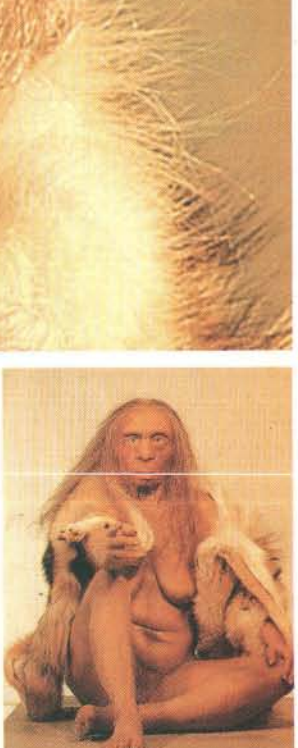
From the tracks alone, we cannot know. For visual interest we opted once more for an adult male and female, the male with one arm draped over the female's shoulder. Perhaps that pose is too anthropomorphic for some tastes, but the gifted English sculptor John Holmes produced such a vivacious result for us in his finished Laetoli figures that we frankly didn't care.

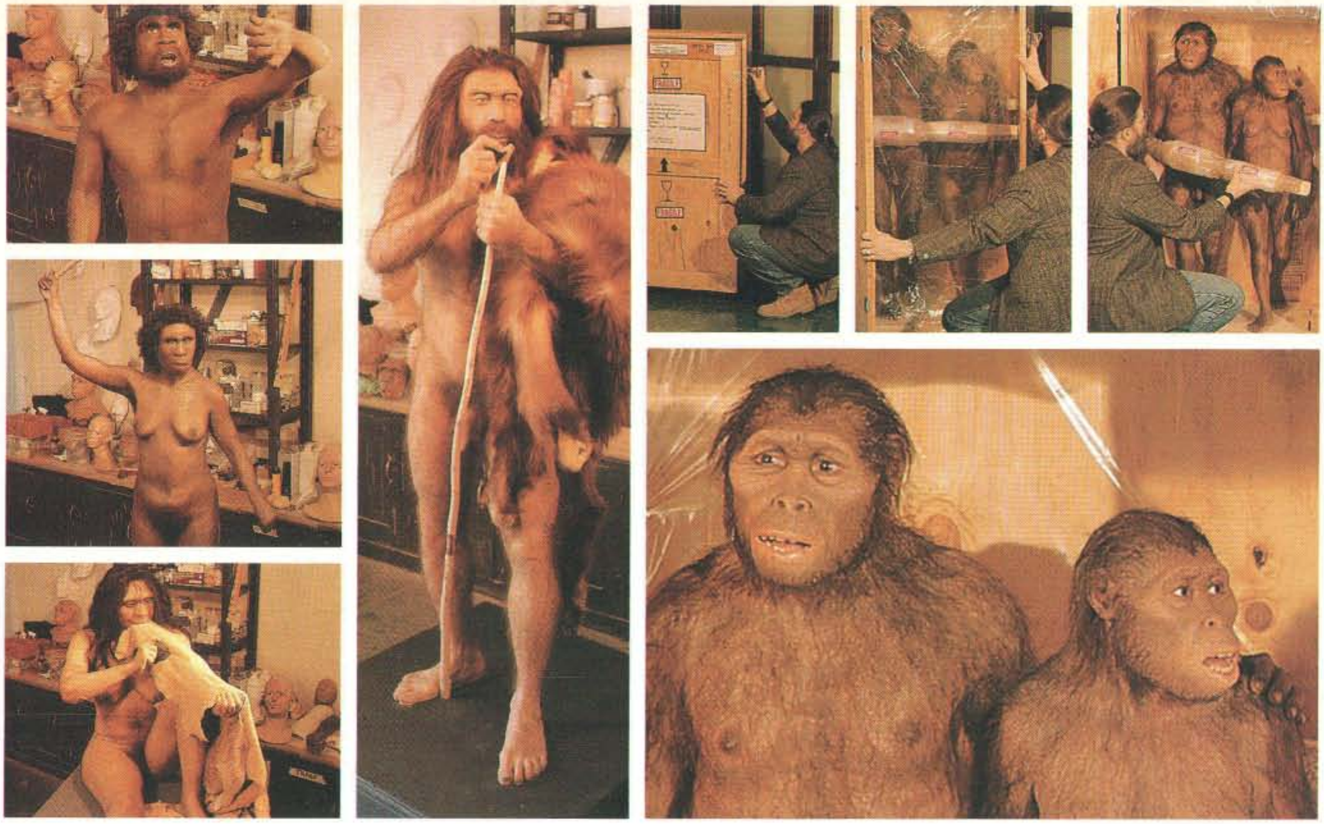
No complete fossil skull of *A. afarensis* has been found.

TO FINISH THE FIGURE, artists insert eyes and painstakingly airbrush lifelike shading on the molded surface. Hairs are placed one by one on the head and body. The nails of the hands and feet are roughened by the artists for an appropriately ragged look.









Composite reconstructions based on skull fragments reveal that it had a rather chimpanzeelike head with a small braincase and a large face. The soft tissue features, of course, posed the usual problems in even more acute form: we had to decide, for example, whether to give the figures a chimpanzeelike nose or something more human.

Moreover, uncertainty about the body structure of these early humans was also a factor. From the known fossils we could infer that they had shorter legs and longer feet than our own. That fact could imply that their gait differed from that of modern people. Yet such an idea conflicts with the evidence of the Laetoli tracks. Caught between the fossils and the humanlike prints, we opted to favor the latter—but that is not the kind of decision that most scientists are comfortable making.

The skin color and body hair of the figures were also highly conjectural. *A. afarensis* lived in the tropics of East Africa well after the ancient forest began to give way to open grassland. This bipedal hominoid probably exploited the relatively rich zone where the forest and savanna intersected. Its skin might have been deeply pigmented for protection against the intense solar radiation of the open country, but if *A. afarensis* spent much of its life in the shade of trees such an adaptation might have been less necessary. It might have had the dense coat of hair typical of forest primates. Alternatively, like our own savanna-adapted species, it might have swapped its coat for a mechanism that shed body heat through evaporation of sweat.

COMPLETED NEANDERTHAL WOMAN, dressed in fur, is one of several extinct human figures that have been reconstructed. Others in the exhibit (*above*) include a younger female and a male Neanderthal (*light-skinned figures*) and a pair representative of *Homo erectus* (*dark-skinned figures*). Curator Willard Whitson uncrates the *Australopithecus afarensis* figures made in England.

There is simply no way to be certain. The *A. afarensis* figures we created have a darkish skin and a fairly sparse covering of hair, but I am sure that many of my colleagues will find good reasons to disagree with our choices.

We had a similarly hard time deciding how closely our male and female *A. afarensis* should resemble each other. Chimpanzees do not show major sex differences in their facial hair, and neither do our figures—partly because we felt that the moustache with which the male was initially endowed made him look a little too much like a Lothario from a 1920s movie. Once again, the decision is a hard one to justify on strictly scientific grounds.

You might conclude that the cumulative result of these unscientific decisions would be purely fantastic figures. Not so. Although I would not stake my life on many of the details I have mentioned, through careful sculpture and respect for the measured skull and body proportions, we have produced evocations of these vanished humans that bring them to life without sacrificing reasonable scientific accuracy. Moreover, while visitors to the American Museum of Natural History are looking at the dioramas, they will also have before them replicas of the actual fossils on which these recreations are based—the best of both worlds!

FURTHER READING

SKULL 1470. Richard E. Leakey in *National Geographic,* Vol. 143, No. 6, pages 819–829; June 1973.

FACING THE PAST. Boyce Rensberger in *Science '81*, Vol. 2, No. 8, pages 40–51; October 1981.

FLESHING OUT THE PAST: RECONSTRUCTING FOSSIL FACES. Michael Anderson in *Discovery* (Yale Peabody Museum of Natural History), Vol. 22, No. 1, pages 11–15; 1990.

ALMOST HUMAN. Tom Waters in *Discover,* Vol. 11, No. 5, pages 42–53; May 1990.

RE-CREATING OUR ANCESTORS. John Neary in *Archaeology,* Vol. 44, No. 1, pages 64–67; January/February 1991.

Paradoxes of Musical Pitch

Certain series of tones appear to ascend or descend infinitely in pitch. Other patterns change when shifted in key and indicate an influence of speech on the perception of music

by Diana Deutsch

The endless staircase stands as a classic visual paradox, continuously tricking the eye into a geometrically impossible journey. First devised in the 1950s by L. S. Penrose and Roger Penrose of the University of London and later made famous by the Dutch artist M. C. Escher, this paradox has a rich set of acoustic counterparts. In the early 1960s Roger N. Shepard of Bell Telephone Laboratories produced a rather remarkable example. He repeatedly played the same sequence of computer-generated tones that moved up in an octave. Instead of hearing the pattern stop and then start again, listeners heard the pattern ascend endlessly in pitch. When Shepard reversed the direction, the subjects heard the pattern descend endlessly.

Such inquiries amount to far more than an acoustic form of aesthetic diversion. Research into the way individuals hear particular sequences of tones reveals how the brain uses different cues to make sense of ambiguous sounds. Indeed, the latest studies suggest that perception of certain musical paradoxes is related to the processing of speech. It appears that during childhood individuals gradually acquire a representation of pitch that is peculiar to a particular language or dialect. Hence, a native of California will perceive a certain pattern of tones differently from a native of the south of England. Such studies have revealed that a common influence on the perception of both speech and musical pitch exists in individuals.

Furthermore, the research has overturned some long-standing assumptions, particularly one concerning perceptual equivalence for musical patterns. This assumption states that a musical passage remains identifiable even if it is presented in a key different from that in which it was originally heard. But on the contrary, certain pitch paradoxes show that this principle is not universal. Rather the brain may completely reinterpret the relations between tones transposed to another key. This notion is as paradoxical as the idea that a visual shape might undergo a metamorphosis if shifted to a different location in space.

Research that supports these and other conclusions has deep historical roots in studies of the physics of music and sound. Indeed, the physical basis of musical pitch has fascinated scientists since antiquity. Pythagoras established that the pitch of a vibrating string varies inversely with its length: the shorter the string, the higher the pitch. He also demonstrated that certain musical intervals—the pitch relation between two tones—correspond to ratios formed by different lengths of string. In the 17th century Galileo and the French mathematician and theologian Marin Mersenne showed that the basis of these associations lay in the relation between string length and frequency of vibration.

Mersenne also demonstrated the existence of overtones, or harmonics, in vibrating bodies. That is, a vibration occurs both at the frequency corresponding to the perceived pitch (the fundamental frequency) and at frequencies that are whole number multiples of the fundamental (harmonics). In other words, a tone whose fundamental frequency is 100 hertz contains components at 200 hertz, 300 hertz, 400 hertz and so on. In the 1930s Jan Schouten of Philips Laboratory in Eindhoven showed that the auditory system exploits this phenomenon. When presented with a harmonic series, we can perceive a pitch that corresponds to the fundamental frequency, even if the fundamental itself is missing.

The relations between pitches enable us to hear musical patterns. When two tones are presented simultaneously or in succession, we perceive a musical interval. Intervals are heard to be the same in size when the fundamental frequencies of their component tones stand in the same ratio. (Technically, the tones within each pair are separated by the same distance in log frequency.)

This principle forms one of the bases of the traditional musical scale. The smallest unit of this scale is the semitone, which is the pitch relation formed by two adjacent notes on a keyboard. The semitone corresponds to a frequency ratio of approximately 18:17. Intervals composed of the same number of semitones are given the same name. For example, the interval corresponding to a ratio of 2:1 (12 semitones) is termed an octave, the interval corresponding to a ratio of 3:2 (seven semitones) is termed a fifth and the ratio 4:3 (five semitones) is called a fourth.

Tones related by octaves are in a sense perceptually equivalent. Each of the 12 semitones in an octave is assigned a name (C, C#, D and so on). The entire scale (called the chromatic scale) consists of the repetitive occurrence of this series of note names across octaves. The note names are identified by subscripts. For example, middle C can be written as C_4. The C one octave lower is C_3, and the one above is C_5.

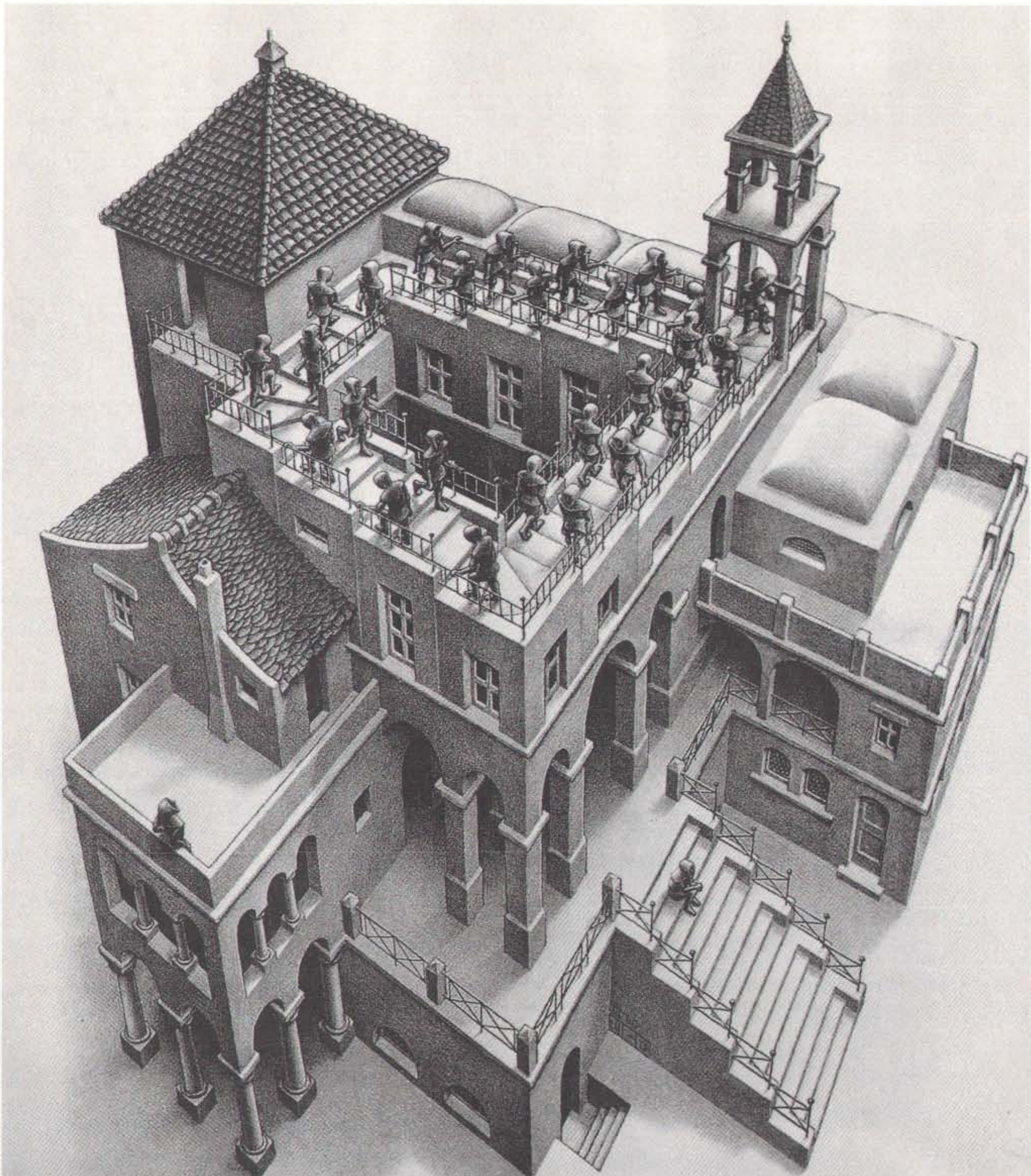
The pitch of a tone can thus be regarded as varying along two dimensions. The first, known as pitch height,

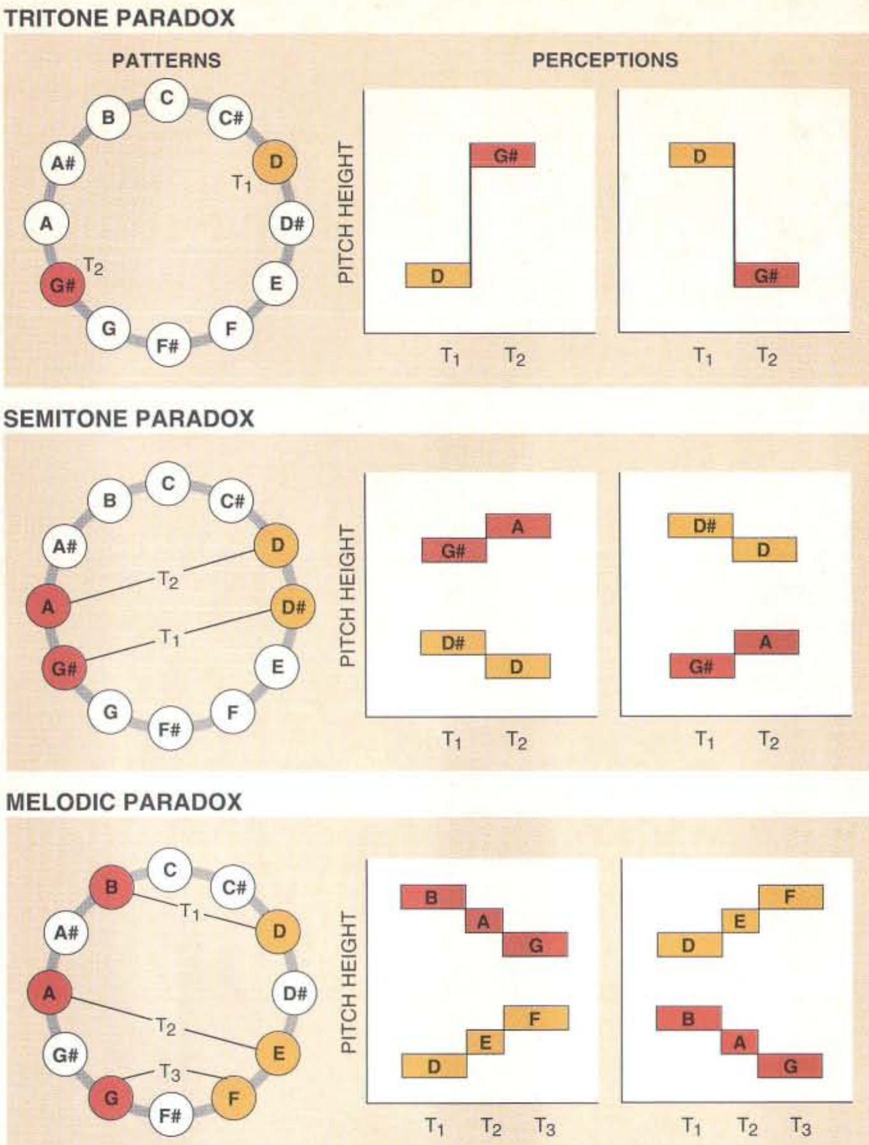
DIANA DEUTSCH is professor of psychology at the University of California, San Diego, where she received her Ph.D. She obtained a first-class honors B.A. degree from the University of Oxford. Her work has centered on the perception of sound and music, including ways in which sound patterns are analyzed and represented in memory, as well as the neurological underpinnings of these processes. This is her second article for *Scientific American*.

***ASCENDING AND DESCENDING*, a lithograph by M. C. Escher, visually parallels the musical illusion of tones that appear to rise or fall endlessly in pitch.**

extends from low to high, which we can experience by sweeping a hand all the way up a keyboard. The second is the circular dimension of pitch class, which defines a tone's position within the octave. Researchers refer to this dimension as the pitch class circle. The circle leads to an immediate assumption: it is nonsensical to ask whether one tone, say, C, is higher than another, such as F#. To clarify the question, one would need to give the octaves in which the two tones occur.

In the absence of such information the human brain still tries to organize tones so that it can judge relative pitch. Shepard demonstrated this phenomenon in 1964. Using a music synthesis program developed by his colleague Max Mathews, he generated a series of tones that were clearly defined in terms of pitch class but in which the octave containing the tones was unclear. Each tone consisted of a set of si-





MUSICAL PARADOXES occur when sequences of tones appear to rise or fall even though the tones lack the physical cues normally used to judge pitch height. Such paradoxes can be understood in terms of the pitch class circle, which represents the tones in an octave. The tones played are opposite one another along the circle. In an example of the phenomenon, called the tritone paradox, D is played at time 1 (T_1), followed by G# at time 2 (T_2). Some listeners heard the sequence ascend; others heard it descend. In a variation called the semitone paradox, D# and G# are presented simultaneously, followed by D and A. Another version, the melodic paradox, uses three pairs of tones. In these cases, some subjects heard the ascending sequence as higher than the descending one, and others heard it as lower. The results show that the subjects must have preferred orientations of the pitch class circle with respect to pitch height.

nusoidal components (smoothly oscillating waves) separated by octaves, so that the tones were composed only of harmonics in the same pitch class.

Shepard found that when two such tones were played, one after the other, subjects heard either an ascending pattern or a descending one. The perceived direction depended on the distance separating the two tones along the pitch class circle: listeners followed the shorter distance between the tones. For example, subjects heard the pair C#-D as ascending, because the shorter distance here is clockwise. Analogously, the pair A-G# was always heard as descending.

This finding enabled Shepard to produce the striking demonstration described at the beginning of this article. A series of tones that repeatedly traverses the pitch class circle in clockwise steps appears to ascend endlessly in pitch. If the series of tones traverses the circle in counterclockwise steps, it appears to descend infinitely.

Jean-Claude Risset, now at the Laboratory for Mechanics and Acoustics at the CNRS in Marseilles, produced an intriguing variant. He created a single tone that glided around the pitch class circle in a clockwise direction. The tone appeared to ascend endlessly. Counterclockwise movement produced a descending tone. Risset used such a gliding tone when he composed incidental music to Pierre Halet's *Little Boy,* a play that depicts the nightmare of a pilot involved in the destruction of Hiroshima. A tone of continuously descending pitch symbolized the falling of the atomic bomb. Risset has also produced a gliding tone that appears to ascend and to descend endlessly at the same time. He has incorporated such tones in orchestral works, with powerful effect.

In my laboratory I recently created pitch circularity effects using a set of tones, each of which constituted a full harmonic series but in which the relative amplitudes (loudnesses) of the harmonics were such as to generate ambiguities of perceived pitch height. Listeners obtained an impression of a series that ascended infinitely in pitch.

These demonstrations of pitch circularity illustrate that the human mind tends to form linkages between elements that are close together rather than those that are far apart. Analogous phenomena can be found in vision. For example, we tend to group together dots that are next to one another and to perceive movement between neighboring lights turned on and off in succession.

What happens, then, when two tones are related by exactly half an octave, such as C followed by F# or G# by D? The tone pairs are separated by the same distance in either direction along the pitch class circle. Over such an interval, termed a tritone, a listener cannot invoke proximity in making judgments about tone pairs. Will perceptions of relative height, then, be ambiguous, or will some other principle be used to avoid ambiguity?

When I considered this question, it occurred to me that another cue was available to the perceptual system. A listener could establish absolute positions for the tones along the pitch class circle. For instance, we can envision the pitch classes to be the numbers on a clock face. Subjects might orient this clock face so that C is in the 12 o'clock position, C# in the one o'clock position and so on. They would hear C-F# (and

B-F and C#-G) as descending and F#-C (and F-B and G-C#) as ascending.

To examine this hypothesis, I presented subjects with just such tone pairs, which were generated by a computer program developed by F. Richard Moore of the University of California at San Diego. Each tone consisted of six sinusoidal components, all in the same pitch class. Subjects judged whether the tones ascended or descended. Their impressions were plotted as a function of pitch class of the first tone of the pair; hence, the pair C-F# would fall under pitch class C. The results strikingly confirmed the hypothesis. The judgments of most subjects showed orderly relations to the pitch class of the first tone: tones in one region of the circle were heard as higher than those in the opposite region.

A wholly unexpected finding emerged as well. The orientation of the pitch class circle with respect to pitch height varied radically from one subject to another. Some subjects heard, for example, the tone pair D-G# as descending, indicating that they oriented D in the upper half of the pitch circle (between the nine and three o'clock positions) and G# in the lower half. Others, however, heard the pattern ascend; they oriented D in the lower half of the circle and G# in the upper half. Subjects heard the illusion reverse itself when the pattern was transposed along the semitone scale.

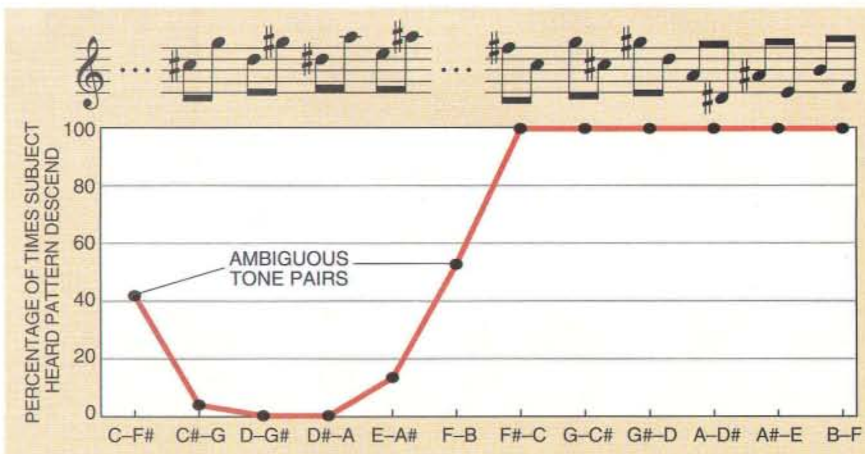
These findings demonstrate that the pitch class circle is not flat with respect to pitch height. Rather, when listeners decide whether pairs of tones related by half an octave form ascending or descending patterns, their judgments are systematically related to the positions of the tones along the pitch class circle. There is, however, striking disagreement as to which region of the circle is tagged as the higher half and which the lower one. These findings, taken together, constitute the tritone paradox.

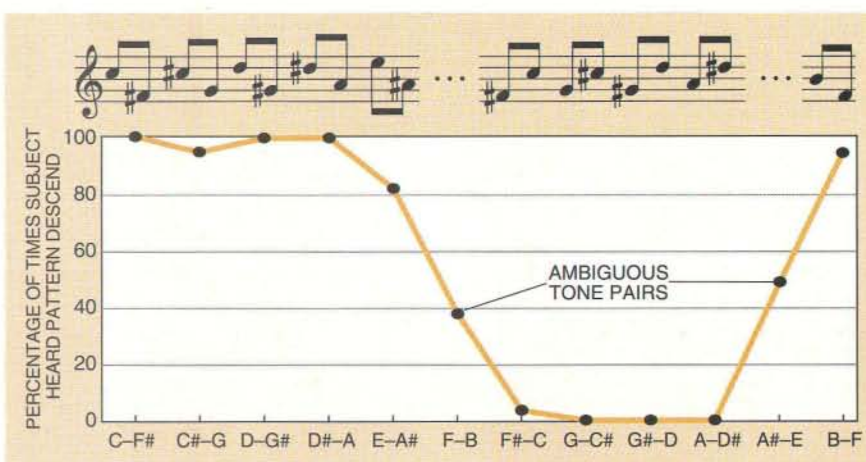
With William L. Kuyper of U.C.S.D. and Yuval Fisher of Cornell University, I performed a large-scale study of the tritone paradox. We selected a group of subjects who were U.C.S.D. undergraduates, had normal hearing and could reliably judge whether pairs of tones formed ascending or descending patterns. We found that the positions of the tones along the pitch class circle strongly influenced individual judgments. The direction of this tendency varied considerably from one subject to another, as expected. Furthermore, computer simulations showed that such perceptions exist to a highly significant extent in the general population.

I conducted other studies to see whether the phenomenon occurred when more complex patterns of tones were used. Specifically, I created the semitone paradox. The pattern consisted of two pairs of tones presented simultaneously; one pair ascended by a semitone, whereas the other pair descended by that interval. The tone pairs were diametrically opposed along the pitch class circle, so that again proximity could not be invoked in making judgments of relative height. Listeners generally perceived this pattern as two stepwise sequences that moved in opposite directions. Some listeners, however, heard the higher line of tones as ascending and the lower line as descending. Others perceived the reverse pattern.

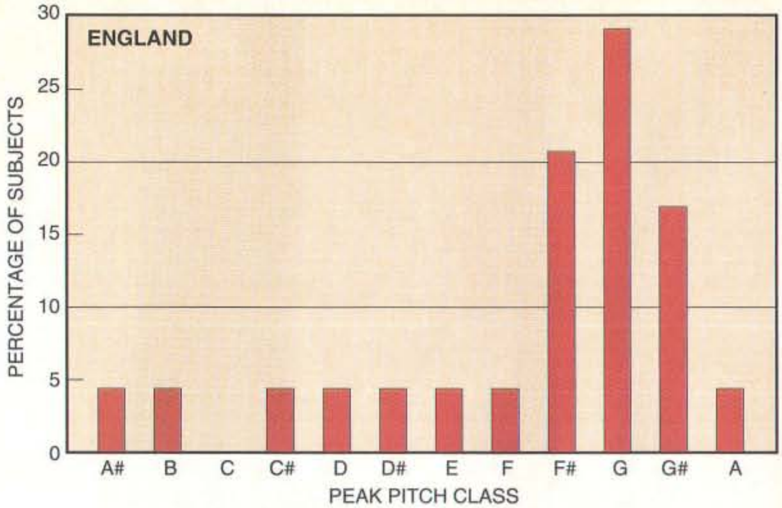
Just as with the tritone paradox, judgments of the semitone paradox reflected an orderly relation between the perceived heights of the tones and their positions along the pitch class circle. In addition, the form of this relation varied substantially across subjects. For example, one subject heard the tones F, F#, G, G#, A and A# as higher than C, C#, D and D#. Yet for a second subject, the tones C#, D and D# were higher than the tones F, F#, G, G#, A, A# and B.

With my colleagues Moore and Mark B. Dolson, I also generated a more complex type of paradox by presenting three simultaneous pairs of tones to our subjects. For example, we played the tones D and B, followed by E and A and then F and G. Listeners generally perceived this pattern as two simultaneous melodies, one higher and one lower, that moved in opposite directions. Some listeners heard the higher melody as descending and the lower melody as ascending; others heard the opposite. When we transposed the pattern by half an octave, the subjects reported that





PERCEPTIONS of the tritone paradox by two subjects differed considerably. The first subject (*top*) clearly heard C#-G, D-G#, D#-A and E-A# as ascending but F#-C, G-C#, G#-D, A-D#, A#-E and B-F as descending. The second one (*bottom*) heard virtually the opposite: pairs heard by the first subject as ascending were perceived to descend, and vice versa. Both found some pairs ambiguous; that is, the tone pairs were heard to ascend and descend about equally often.



BRITISH AND CALIFORNIAN POPULATIONS perceive the tritone paradox in virtually opposite ways. Each graph shows the proportion of subjects that perceived a tone to reside at the peak region of the pitch class circle (that is, at the 12

the higher and lower melodies appeared to exchange positions. This melodic paradox shows that the subjects were perceptually preserving the relative heights of the different pitch classes.

I later presented the melodic paradox in six different keys: C, D, E, F#, G# and A#. These keys correspond to six equal steps along the pitch class circle. The subjects' perceptions of ascending and descending melodies depended on the key in which the tones were played. In another demonstration, I played the pattern as the keys shifted up in whole tone steps, so that the pattern was first in the key of C, then D and so on. Most people first heard the pattern one way; as the keys shifted, the pattern turned upside down and finally righted itself. In other words, the pattern appeared to rotate, in a fashion analogous to the rotation of shapes in vision.

These paradoxes lead to some fascinating conclusions about the auditory perceptual system. Clearly, the principle of equivalence under transposition, which investigators had thought to be universal, can be violated. Shifts in key most definitely affect the perception of certain patterns of tones. Another surprising conclusion concerns the phenomenon of absolute pitch, which is the ability to name a note just from hearing it. Musicians prize this presumed-to-be-rare faculty. Yet the experiments show that the ability is in greater supply than has been thought (at least in a partial form): individuals can perceive notes as higher or lower simply on the basis of pitch class.

The findings support research conducted in the early 1980s by Ernst Terhardt and Manfred Seewann of the Technical University of Munich and W. Dixon Ward of the University of Minnesota. These investigators found that musicians could generally determine whether well-known passages were played in the correct key, even though most of them did not have absolute pitch as conventionally defined. In fact, many succeeded even when the difference was as small as a semitone.

Finally, the paradoxes might have some implications for everyday listening to music. I found that different types of tone complexes—similar to groups of tones produced by natural instruments playing in octave relation—create the same paradoxes as do the more simple tones considered here. Furthermore, the effects endured when the tones were subjected to such time-varying manipulations as rapid fluctuations in pitch (vibratos), quick changes in loudness or fast decays.

Given the variety of sounds that produce these paradoxes, it is likely they also occur in music played by natural instruments. They could lead to subtle perceptual differences that could be of aesthetic importance in some pieces, especially orchestral works in which the composer has created an impression of ambiguity. Such ambiguities appear in Debussy's *Nocturnes.*

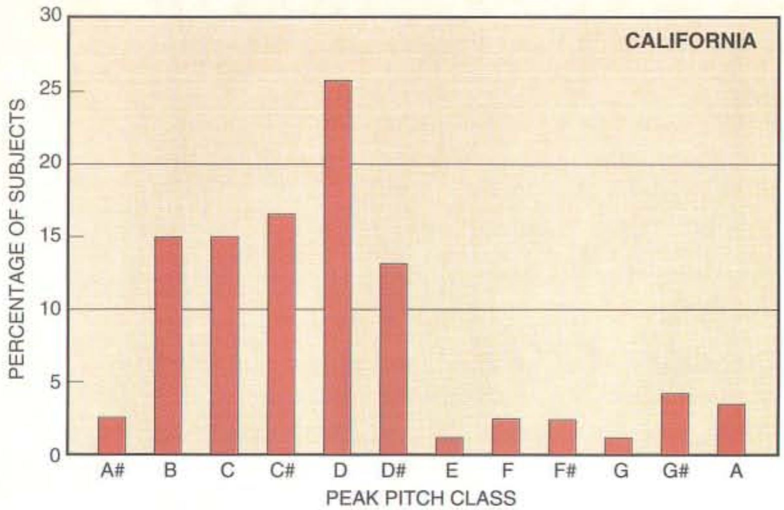
What can be the basis of this unexpected relation between pitch class and perceived height and of the individual differences in the manifestation of these paradoxes? The experiments conducted on the general population, where no correlation with musical training was found, indicate that the paradoxes are not musical in origin. In other studies, I ruled out simple characteristics of the hearing mechanism. For instance, I conducted experiments in which the odd-numbered components of each tone were played to one ear and the even-numbered ones to the other. The tritone paradox persisted, indicating that the brain produces the paradox by integrating information from both ears.

Based on these conclusions, and several informal observations, I conjectured that speech patterns might be responsible for the results. I had noticed that people from other countries often heard the paradox differently than did Californians. For example, individuals from the south of England tended to hear the tritone paradox in an opposite manner to that typical of subjects born in California: it appeared that what one group heard as ascending, the other heard as descending.

What might account for this observation? I hypothesized that people might have a long-term representation of the pitch range of their speaking voices. Included in this representation might be a preference for the octave band that contains the largest proportion of pitch values in their speech. I further hypothesized that the listener fixes the pitch classes defining this octave band at the highest positions along the pitch class circle (that is, near the 12 o'clock position). This definition determines the way the listener hears the paradoxes.

To examine this hypothesis, I undertook a study together with my colleagues Tom North and Lee Ray. We selected subjects whose judgments of the tritone paradox showed clear relations between pitch class and perceived height. We then recorded 15 minutes of spontaneous speech from each subject. From this recording, we determined the octave band containing the the largest number of pitch values in his or her speech.

Comparing the results from each subject, we found a significant correspondence between the pitch classes



o'clock position). In general, when one group heard a pattern descend, the other heard it ascend. It remains to be seen whether such differences have implications for the appreciation or performance of music.

defining this octave band for speech and those defining the highest position along the pitch class circle, as determined by each individual's judgments of the tritone paradox. The findings from this experiment substantiate the hypothesis that perception of the tritone paradox is based on the listener's representation of the pitch class circle—that is, on a kind of perceptual template. The orientation of this template is related to the pitch range of one's speaking voice.

How do these preferences arise? One interpretation would suggest that the listener's vocal range is completely innate. Indeed, some auditory illusions appear to result from differences at a basic neurological level. Research has shown, for instance, that the perception of some illusions correlates with the handedness of the listener [see "Musical Illusions," by Diana Deutsch; SCIENTIFIC AMERICAN, October 1975]. Another possibility is that a speech template could be acquired developmentally, through exposure to speech produced by others. Individuals would use this template to constrain their own speech output and to evaluate perceived speech.

The characteristics of such a template would therefore be expected to vary for those who speak in different languages or dialects, in a fashion similar to such other speech features as vowel quality. Following this line of reasoning, the orientation of the pitch class circle with respect to height should be similar for individuals who speak in the same language or dialect but should vary for those who speak in different languages or dialects.

The study of California undergraduates conducted by Kuyper, Fisher and me provided the initial evidence for the developmental acquisition of a speech template. Although no information was obtained concerning the linguistic backgrounds of these subjects, the majority had probably grown up in California and were from the same linguistic subculture. An orderly distribution of peak pitch class emerged among these subjects: C# and D occurred most frequently as peak pitch classes, followed by C and D#.

Given this preliminary finding, I carried out a more detailed study. I chose two groups for this purpose. The first consisted of 24 individuals who had grown up in California and spoke with the regional accent. The second was made up of 12 people from the south of England. In the experiment, I found that in the English group, F#, G and G# occurred most often as peak pitch classes. But in the California group, B, C, C#, D and D# occurred most frequently. Musical training appeared to have no effect, and neither did age or gender. This experiment provides strong support for the view that, through a learning process, the individual acquires a representation of the pitch class circle that has a particular orientation with respect to height.

It seems safe to assume that we employ such a template in our own speech and in the interpretation of the speech of others. Thus, a template based on pitch class rather than pitch height can be invoked by male and female speakers, even though they speak in different pitch ranges.

Some evolutionary value for such a template must exist. It could serve to provide a framework, common to a particular dialect, within which the pitch of a speaker's voice may be evaluated for emotion. A listener may also invoke the template in the communication of syntactic elements of speech.

It is likely that other disagreements in pitch perception await discovery. Unfortunately, we do not describe our musical experiences in terms that are precise or accurate enough for such differences to become apparent from informal observation. Only in the laboratory can researchers develop a clear idea of what a listener actually perceives and further define relations between speech and music. In many ways the studies confirm what some philosophers and musicians have argued for centuries: that one can achieve expressiveness in music by incorporating characteristics of the speaking voice, such as tempo and pitch range and variability. As the composer Mussorgsky wrote in his autobiography, "the function of art [is] the reproduction in musical sounds not merely of feelings, but first and foremost of human speech."

FURTHER READING

CIRCULARITY IN JUDGMENTS OF RELATIVE PITCH. Roger N. Shepard in *Journal of the Acoustical Society of America,* Vol. 36, No. 12, pages 2346–2353; December 1964.

AUDITORY PATTERN RECOGNITION. Diana Deutsch in *Handbook of Perception and Human Performance.* Edited by Kenneth R. Boff, Lloyd Kaufman and James P. Thomas. Wiley Interscience, 1986.

THE TRITONE PARADOX: AN INFLUENCE OF LANGUAGE ON MUSIC PERCEPTION. Diana Deutsch in *Music Perception,* Vol. 8, No. 4, pages 335–347; Summer 1991.

SOME NEW PITCH PARADOXES AND THEIR IMPLICATIONS. Diana Deutsch in *Philosophical Transactions of the Royal Society of London,* Series B, Vol. 336, No. 1278, pages 391–397; June 1992.

Achieving Electronic Privacy

A cryptographic invention known as a blind signature permits numbers to serve as electronic cash or to replace conventional identification. The author hopes it may return control of personal information to the individual

by David Chaum

Every time you make a telephone call, purchase goods using a credit card, subscribe to a magazine or pay your taxes, that information goes into a data base somewhere. Furthermore, all these records can be linked so that they constitute in effect a single dossier on your life—not only your medical and financial history but also what you buy, where you travel and whom you communicate with. It is almost impossible to learn the full extent of the files that various organizations keep on you, much less to assure their accuracy or to control who may gain access to them.

Organizations link records from different sources for their own protection. Certainly it is in the interest of a bank looking at a loan application to know that John Doe has defaulted on four similar loans in the past two years. The bank's possession of that information also helps its other customers, to whom the bank passes on the cost of bad loans. In addition, these records permit Jane Roe, whose payment history is impeccable, to establish a charge account at a shop that has never seen her before.

That same information in the wrong hands, however, provides neither protection for businesses nor better service for consumers. Thieves routinely use a stolen credit card number to trade on their victims' good payment records; murderers have tracked down their targets by consulting government-maintained address records. On another level, the U.S. Internal Revenue Service has attempted to single out taxpayers for audits based on estimates of household income compiled by mailing-list companies.

The growing amounts of information that different organizations collect about a person can be linked because all of them use the same key—in the U.S. the social security number—to identify the individual in question. This identifier-based approach perforce trades off security against individual liberties. The more information that organizations have (whether the intent is to protect them from fraud or simply to target marketing efforts), the less privacy and control people retain.

Over the past eight years, my colleagues and I at CWI (the Dutch nationally funded Center for Mathematics and Computer Science in Amsterdam) have developed a new approach, based on fundamental theoretical and practical advances in cryptography, that makes this trade-off unnecessary. Transactions employing these techniques avoid the possibility of fraud while maintaining the privacy of those who use them.

In our system, people would in effect give a different (but definitively verifiable) pseudonym to every organization they do business with and so make dossiers impossible. They could pay for goods in untraceable electronic cash or present digital credentials that serve the function of a banking passbook, driver's license or voter registration card without revealing their identity. At the same time, organizations would benefit from increased security and lower record-keeping costs.

Recent innovations in microelectronics make this vision practical by providing personal "representatives" that store and manage their owners' pseudonyms, credentials and cash. Microprocessors capable of carrying out the necessary algorithms have already been embedded in pocket computers the size and thickness of a credit card. Such systems have been tested on a small scale and could be in widespread use by the middle of this decade.

The starting point for this approach is the digital signature, first proposed in 1976 by Whitfield Diffie, then at Stanford University. A digital signature transforms the message that is signed so that anyone who reads it can be sure of who sent it [see "The Mathematics of Public-Key Cryptography," by Martin E. Hellman; SCIENTIFIC AMERICAN, August 1979]. These signatures employ a secret key used to sign messages and a public one used to verify them. Only a message signed with the private key can be verified by means of the public one. Thus, if Alice wants to send a signed message to Bob (these two are the cryptographic community's favorite hypothetical characters), she transforms it using her private key, and he applies her public key to make sure that it was she who sent it. The best methods known for producing forged signatures would require many years, even using computers billions of times faster than those now available.

To see how digital signatures can provide all manner of unforgeable credentials and other services, consider how they might be used to provide an electronic replacement for cash. The First Digital Bank would offer electronic bank notes: messages signed using a particular private key. All messages bearing one key might be worth a dollar, all those bearing a different key five dollars, and so on for whatever denominations were needed. These electronic bank notes could be authenticated using the corresponding public key, which the bank has made a matter of record. First Digital would also make public a key to authenticate electronic documents

DAVID CHAUM is head of the Cryptography Group at the Center for Mathematics and Computer Science (CWI) in Amsterdam. He is also a founder of DigiCash, which develops electronic payment systems. Chaum received his Ph.D. in computer science from the University of California, Berkeley, in 1982 and joined CWI in 1984. He helped to found the International Association for Cryptologic Research and remains active on its board; he also consults internationally on cryptology.

sent from the bank to its customers.

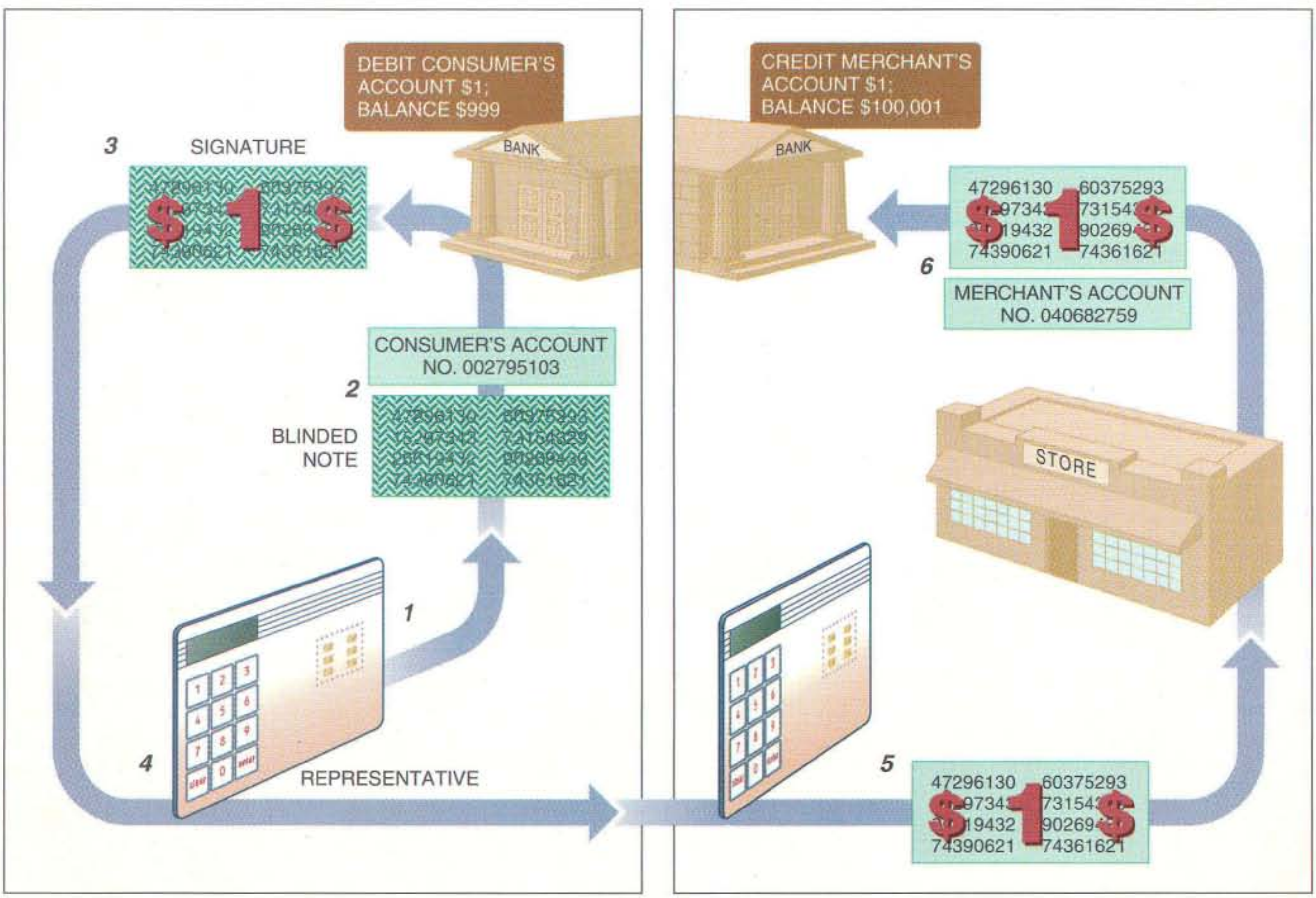
To withdraw a dollar from the bank, Alice generates a note number (each note bears a different number, akin to the serial number on a bill); she chooses a 100-digit number at random so that the chance anyone else would generate the same one is negligible. She signs the number with the private key corresponding to her "digital pseudonym" (the public key that she has previously established for use with her account). The bank verifies Alice's signature and removes it from the note number, signs the note number with its worth-one-dollar signature and debits her account. It then returns the signed note along with a digitally signed withdrawal receipt for Alice's records. In practice, the creation, signing and transfer of note numbers would be carried out by Alice's card computer. The power of the cryptographic protocols, however, lies in the fact that they are secure regardless of physical medium: the same transactions could be carried out using only pencil and paper.

When Alice wants to pay for a purchase at Bob's shop, she connects her "smart" card with his card reader and transfers one of the signed note numbers the bank has given her. After verifying the bank's digital signature, Bob transmits the note to the bank, much as a merchant verifies a credit card transaction today. The bank reverifies its signature, checks the note against a list of those already spent and credits Bob's account. It then transmits a "deposit slip," once again unforgeably signed with the appropriate key. Bob hands the merchandise to Alice along with his own digitally signed receipt, completing the transaction.

This system provides security for all three parties. The signatures at each stage prevent any one from cheating either of the others: the shop cannot deny that it received payment, the bank cannot deny that it issued the notes or that it accepted them from the shop for deposit, and the customer can neither deny withdrawing the notes from her account nor spend them twice.

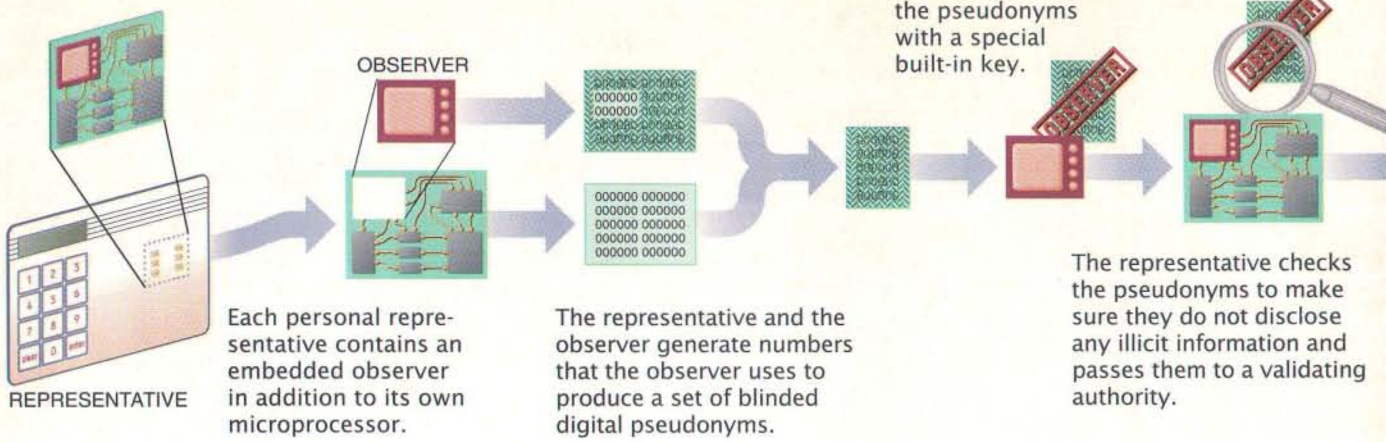
This system is secure, but it has no privacy. If the bank keeps track of note numbers, it can link each shop's deposit to the corresponding withdrawal and so determine precisely where and when Alice (or any other account holder) spends her money. The resulting dossier is far more intrusive than those now being compiled. Furthermore, records based on digital signatures are more vulnerable to abuse than conventional files. Not only are they self-authenticating (even if they are copied, the information they contain can be verified by anyone), but they also permit a person who has a particular kind of information to prove its existence without either giving the information away or revealing its source. For example, someone might be able to prove incontrovertibly that Bob had telephoned Alice on 12 separate occasions without having to reveal the time and place of any of the calls.

I have developed an extension of digital signatures, called blind signatures, that can restore privacy. Before send-



DIGITAL CASH flows tracelessly from bank through consumer and merchant before returning to the bank. Using a small computer "representative," a person creates a random number to serve as a bank note. The bank debits the appropriate account and signs the note with an unforgeable digital signature indicating its value. The bank credits the merchant's account when the note is presented for payment. A technique known as a blind signature prevents the bank from seeing the note number so the bank will be unable to correlate withdrawals from one account with deposits to another.

How to Create Secure Digital Pseudonyms



ing a note number to the bank for signing, Alice in essence multiplies it by a random factor. Consequently, the bank knows nothing about what it is signing except that it carries Alice's digital signature. After receiving the blinded note signed by the bank, Alice divides out the blinding factor and uses the note as before.

The blinded note numbers are "unconditionally untraceable"—that is, even if the shop and the bank collude, they cannot determine who spent which notes. Because the bank has no idea of the blinding factor, it has no way of linking the note numbers that Bob deposits with Alice's withdrawals. Whereas the security of digital signatures is dependent on the difficulty of particular computations, the anonymity of blinded notes is limited only by the unpredictability of Alice's random numbers. If she wishes, however, Alice can reveal these numbers and permit the notes to be stopped or traced.

Blinded electronic bank notes protect an individual's privacy, but because each note is simply a number, it can be copied easily. To prevent double spending, each note must be checked on-line against a central list when it is spent. Such a verification procedure might be acceptable when large amounts of money are at stake, but it is far too expensive to use when someone is just buying a newspaper. To solve this problem, my colleagues Amos Fiat and Moni Naor and I have proposed a method for generating blinded notes that requires the payer to answer a random numeric query about each note when making a payment. Spending such a note once does not compromise unconditional untraceability, but spending it twice reveals enough information to make the payer's account easily traceable. In fact, it can yield a digitally signed confession that cannot be forged even by the bank.

Cards capable of such anonymous payments already exist. Indeed, DigiCash, a company with which I am associated, has installed equipment in two office buildings in Amsterdam that permits copiers, fax machines, cafeteria cash registers and even coffee vending machines to accept digital "bank notes." We have also demonstrated a system for automatic toll collection in which automobiles carry a card that responds to radioed requests for payment even as they are traveling at highway speeds.

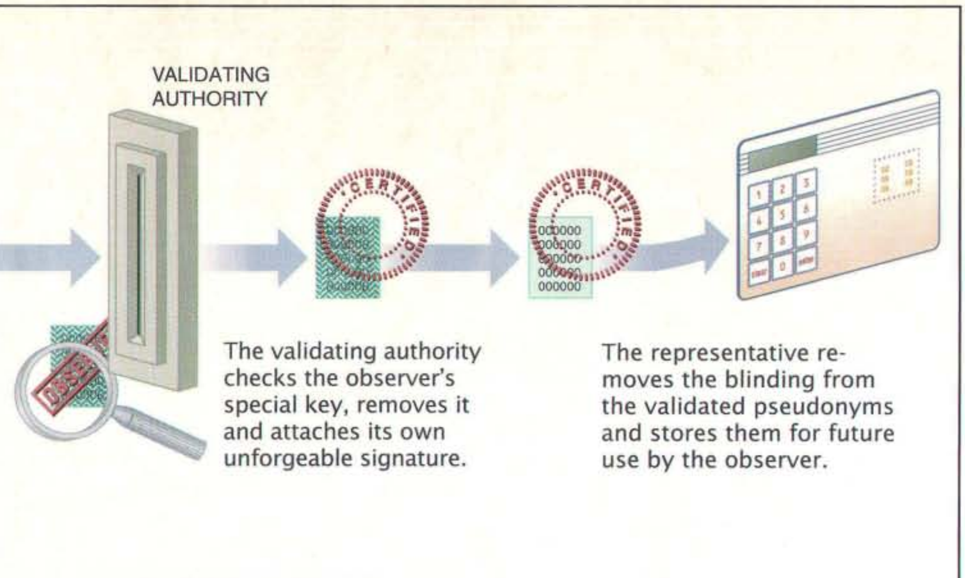
My colleagues and I call a computer that handles such cryptographic transactions a "representative." A person might use different computers as representatives depending on which was convenient: Bob might purchase software (transmitted to him over a network) by using his home computer to produce the requisite digital signatures, go shopping with a "palm-top" personal computer and carry a smart credit card to the beach to pay for a drink or crab cakes. Any of these machines could represent Bob in a transaction as long as the digital signatures each generates are under his control.

Indeed, such computers can act as representatives for their owners in virtually any kind of transaction. Bob can trust his representative and Alice hers because they have each chosen their own machine and can reprogram it at will (or, in principle, build it from scratch). Organizations are protected by the cryptographic protocol and so do not have to trust the representatives.

The prototypical representative is a smart credit-card-size computer containing memory and a microprocessor. It also incorporates its own keypad and display so that its owner can control the data that are stored and exchanged. If a shop provided the keypad and display, it could intercept passwords on their way to the card or show one price to the customer and another to the card. Ideally, the card would communicate with terminals in banks and shops by a short-range communications link such as an infrared transceiver and so need never leave its owner's hands.

When asked to make a payment, the representative would present a summary of the particulars and await approval before releasing funds. It would also insist on electronic receipts from organizations at each stage of all transactions to substantiate its owner's position in case of dispute. By requiring a password akin to the PIN (personal identifying number) now used for bank cards, the representative could safeguard itself from abuse by thieves. Indeed, most people would probably keep backup copies of their keys, electronic bank notes and other data; they could recover their funds if a representative were lost or stolen.

Personal representatives offer excellent protection for individual privacy, but organizations might prefer a mechanism to protect their interests as strongly as possible. For example, a bank might want to prevent double spending of bank notes altogether rather than



leak any information and passes it to Bob. The observer is programmed to sign only one such statement for any given note.

Many transactions do not simply require a transfer of money. Instead they involve credentials—information about an individual's relationship to some organization. In today's identifier-based world, all of a person's credentials are easily linked. If Alice is deciding whether to sell Bob insurance, for example, she can use his name and date of birth to gain access to his credit status, medical records, motor vehicle file and criminal record, if any.

Using a representative, however, Bob would establish relationships with different organizations under different digital pseudonyms. Each of them can recognize him unambiguously, but none of their records can be linked.

simply detecting it after the fact. Some organizations might also want to ensure that certain digital signatures are not copied and widely disseminated (even though the copying could be detected afterward).

Organizations have already begun issuing tamperproof cards (in effect, their own representatives) programmed to prevent undesirable behavior. But these cards can act as "Little Brothers" in everyone's pocket.

We have developed a system that satisfies both sides. An observer—a tamper-resistant computer chip, issued by some entity that organizations can trust—acts like a notary and certifies the behavior of a representative in which it is embedded. Philips Industries has recently introduced a tamper-resistant chip that has enough computing power to generate and verify digital signatures. Since then, Siemens, Thomson CSF and Motorola have announced plans for similar circuits, any of which could easily serve as an observer.

The central idea behind the protocol for observers is that the observer does not trust the representative in which it resides, nor does the representative trust the observer. Indeed, the representative must be able to control all data passing to or from the observer; otherwise the tamperproof chip might be able to leak information to the world at large.

When Alice first acquires an observer, she places it in her smart-card representative and takes it to a validating authority. The observer generates a batch of public and private key pairs from a combination of its own random numbers and numbers supplied by the card. The observer does not reveal its numbers but reveals enough information about them so that the card can later check whether its numbers were in fact used to produce the resulting keys. The card also produces random data that the observer will use to blind each key.

Then the observer blinds the public keys, signs them with a special built-in key and gives them to the card. The card verifies the blinding and the signature and checks the keys to make sure they were correctly generated. It passes the blinded, signed keys to the validating authority, which recognizes the observer's built-in signature, removes it and signs the blinded keys with its own key. The authority passes the keys back to the card, which unblinds them. These keys, bearing the signature of the validating authority, serve as digital pseudonyms for future transactions; Alice can draw on them as needed.

An observer could easily prevent (rather than merely detect) double spending of electronic bank notes. When Alice withdraws money from her bank, the observer witnesses the process and so knows what notes she received. At Bob's shop, when Alice hands over a note from the bank, she also hands over a digital pseudonym (which she need use only once) signed by the validating authority. Then the observer, using the secret key corresponding to the validated pseudonym, signs a statement certifying that the note will be spent only once, at Bob's shop and at this particular time and date. Alice's card verifies the signed statement to make sure that the observer does not

In order to be of use, a digital credential must serve the same function as a paper-based credential such as a driver's license or a credit report. It must convince someone that the person attached to it stands in a particular relation to some issuing authority. The name, photograph, address, physical description and code number on a driver's license, for example, serve merely to link it to a particular person and to the corresponding record in a data base. Just as a bank can issue unforgeable, untraceable electronic cash, so too could a university issue signed digital diplomas or a credit-reporting bureau issue signatures indicating a person's ability to repay a loan.

When the young Bob graduates with honors in medieval literature, for example, the university registrar gives his representative a digitally signed message asserting his academic credentials. When Bob applies to graduate school, however, he does not show the admissions committee that message. Instead his representative asks its observer to sign a statement that he has a B.A. cum laude and that he qualifies for financial aid based on at least one of the university's criteria (but without revealing which ones). The observer, which has verified and stored each of Bob's credentials as they come in, simply checks its memory and signs the statement if it is true.

In addition to answering just the right question and being more reliable than paper ones, digital credentials would be both easier for individuals to obtain and to show and cheaper for organizations to issue and to authenticate. People would no longer need to fill out long and revealing forms. In-

stead their representatives would convince organizations that they meet particular requirements without disclosing any more than the simple fact of qualification. Because such credentials reveal no unnecessary information, people would be willing to use them even in contexts where they would not willingly show identification, thus enhancing security and giving the organization more useful data than it would otherwise acquire.

Positive credentials, however, are not the only kind that people acquire. They may also acquire negative credentials, which they would prefer to conceal: felony convictions, license suspensions or statements of pending bankruptcy. In many cases, individuals will give organizations the right to inflict negative credentials on them in return for some service. For instance, when Alice borrows books from a library, her observer would be instructed to register an overdue notice unless it had received a receipt for the books' return within some fixed time.

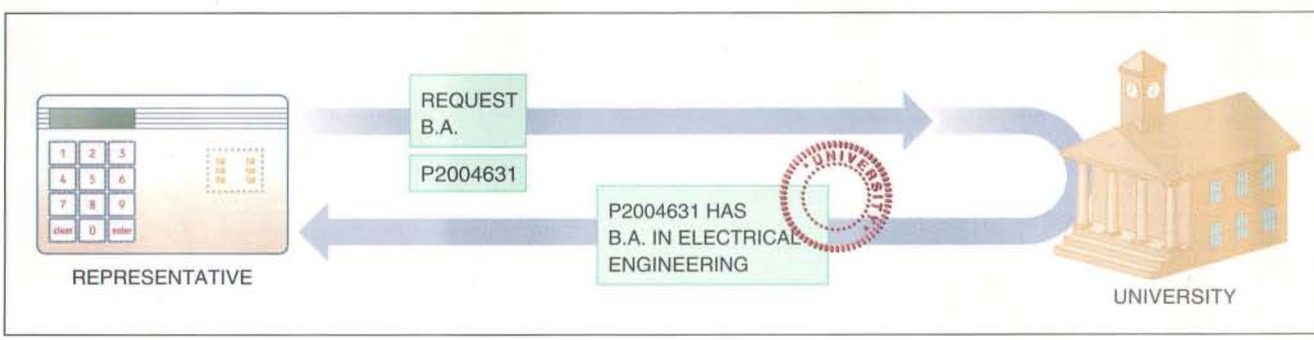
Once the observer has registered a negative credential, an organization can find out about it simply by asking the observer (through the representative) to sign a message attesting to its presence or absence. Although a representative could muzzle the observer, it could not forge an assertion about the state of its credentials. In other cases, organizations might simply take the lack of a positive credential as a negative one. If Bob signs up for skydiving lessons, his instructors may assume that he is medically unfit unless they see a credential to the contrary.

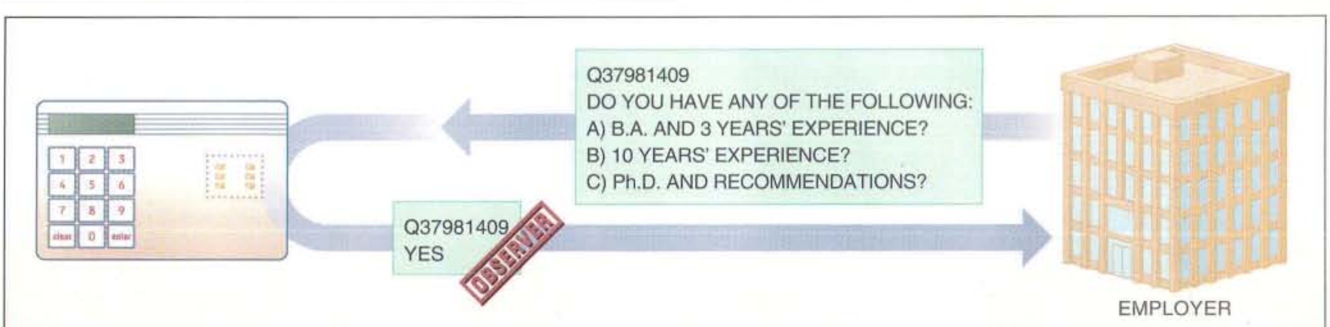
For most credentials, the digital signature of an observer is sufficient to convince anyone of its authenticity. Under some circumstances, however, an organization might insist that an observer demonstrate its physical presence. Otherwise, for example, any number of people might be able to gain access to nontransferable credentials (perhaps a health club membership) by using representatives connected by concealed communications links to another representative containing the desired credential.

Moreover, the observer must carry out this persuasion while its input and output are under the control of the representative that contains it. When Alice arrives at her gym, the card reader at the door sends her observer a series of single-bit challenges. The observer immediately responds to each challenge with a random bit that is encoded by the card on its way back to the organization. The speed of the observer's response establishes that it is inside the card (since processing a single bit introduces almost no delay compared with the time that signals take to traverse a wire). After a few dozen iterations the card reveals to the observer how it encoded the responses; the observer signs a statement including the challenges and encoded responses only if it has been a party to that challenge-response sequence. This process convinces the organization of the observer's presence without allowing the observer to leak information.

Organizations can also issue credentials using methods that depend on cryptography alone rather than on observers. Although currently practical approaches can handle only relatively simple queries, Gilles Brassard of the University of Montreal, Claude Crépeau of the École Normale Supérieure and I have shown how to answer arbitrary combinations of questions about even the most complex credentials while maintaining unconditional unlinkability. The concealment of purely cryptographic negative credentials could be detected by the same kinds of techniques that detect double spending of electronic bank notes. And a combination of these cryptographic methods with observers would offer accountability after the fact even if the observer chip were somehow compromised.

The improved security and privacy of digital pseudonyms exact a price: responsibility. At present, for example, people can disavow credit card purchases made over the tele-





DIGITAL CREDENTIALS put personal information under the control of an individual's representative and its observer. When Alice (one of the author's two hypothetical characters) finishes her undergraduate work, the university gives her a digitally signed degree. Later, her observer can use its knowledge of the degree to answer questions about her qualifications without revealing any more information about her than absolutely necessary.

phone or cash withdrawals from an automatic teller machine (ATM). The burden of proof is on the bank to show that no one else could have made the purchase or withdrawal. If computerized representatives become widespread, owners will establish all their own passwords and so control access to their representatives. They will be unable to disavow a representative's actions.

Current tamper-resistant systems such as ATMs and their associated cards typically rely on weak, inflexible security procedures because they must be used by people who are neither highly competent nor overly concerned about security. If people supply their own representatives, they can program them for varying levels of security as they see fit. (Those who wish to trust their assets to a single four-digit code are free to do so, of course.) Bob might use a short PIN (or none at all) to authorize minor transactions and a longer password for major ones. To protect himself from a robber who might force him to give up his passwords at gunpoint, he could use a "duress code" that would cause the card to appear to operate normally while hiding its more important assets or credentials or perhaps alerting the authorities that it had been stolen.

COMPUTERIZED CREDIT CARD developed by Toshiba and Visa International contains a microprocessor, memory, keypad and display. Although this card identifies its user during transactions, the same hardware could be reprogrammed as a personal representative for spending digital cash.

A personal representative could also recognize its owner by methods that most people would consider unreasonably intrusive in an identifier-based system; a notebook computer, for example, might verify its owner's voice or even fingerprints. A supermarket checkout scanner capable of recognizing a person's thumbprint and debiting the cost of groceries from their savings account is Orwellian at best. In contrast, a smart credit card that knows its owner's touch and doles out electronic bank notes is both anonymous and safer than cash. In addition, incorporating some essential part of such identification technology into the tamperproof observer would make such a card suitable even for very high security applications.

Computerized transactions of all kinds are becoming ever more pervasive. More than half a dozen countries have developed or are testing chip cards that would replace cash. In Denmark, a consortium of banking, utility and transport companies has announced a card that would replace coins and small bills; in France, the telecommunications authorities have proposed general use of the smart cards now used at pay telephones. The government of Singapore has requested bids for a system that would communicate with cars and charge their smart cards as they pass various points on a road (as opposed to the simple vehicle identification systems already in use in the U.S. and elsewhere). And cable and satellite broadcasters are experimenting with smart cards for delivering pay-per-view television. All these systems, however, are based on cards that identify themselves during every transaction.

If the trend toward identifier-based smart cards continues, personal privacy will be increasingly eroded. But in this conflict between organizational security and individual liberty, neither side emerges as a clear winner. Each round of improved identification techniques, sophisticated data analysis or extended linking can be frustrated by widespread noncompliance or even legislated limits, which in turn may engender attempts at further control.

Meanwhile, in a system based on representatives and observers, organizations stand to gain competitive and political advantages from increased public confidence (in addition to the lower costs of pseudonymous record-keeping). And individuals, by maintaining their own cryptographically guaranteed records and making only necessary disclosures, will be able to protect their privacy without infringing on the legitimate needs of those with whom they do business.

The choice between keeping information in the hands of individuals or of organizations is being made each time any government or business decides to automate another set of transactions. In one direction lies unprecedented scrutiny and control of people's lives, in the other, secure parity between individuals and organizations. The shape of society in the next century may depend on which approach predominates.

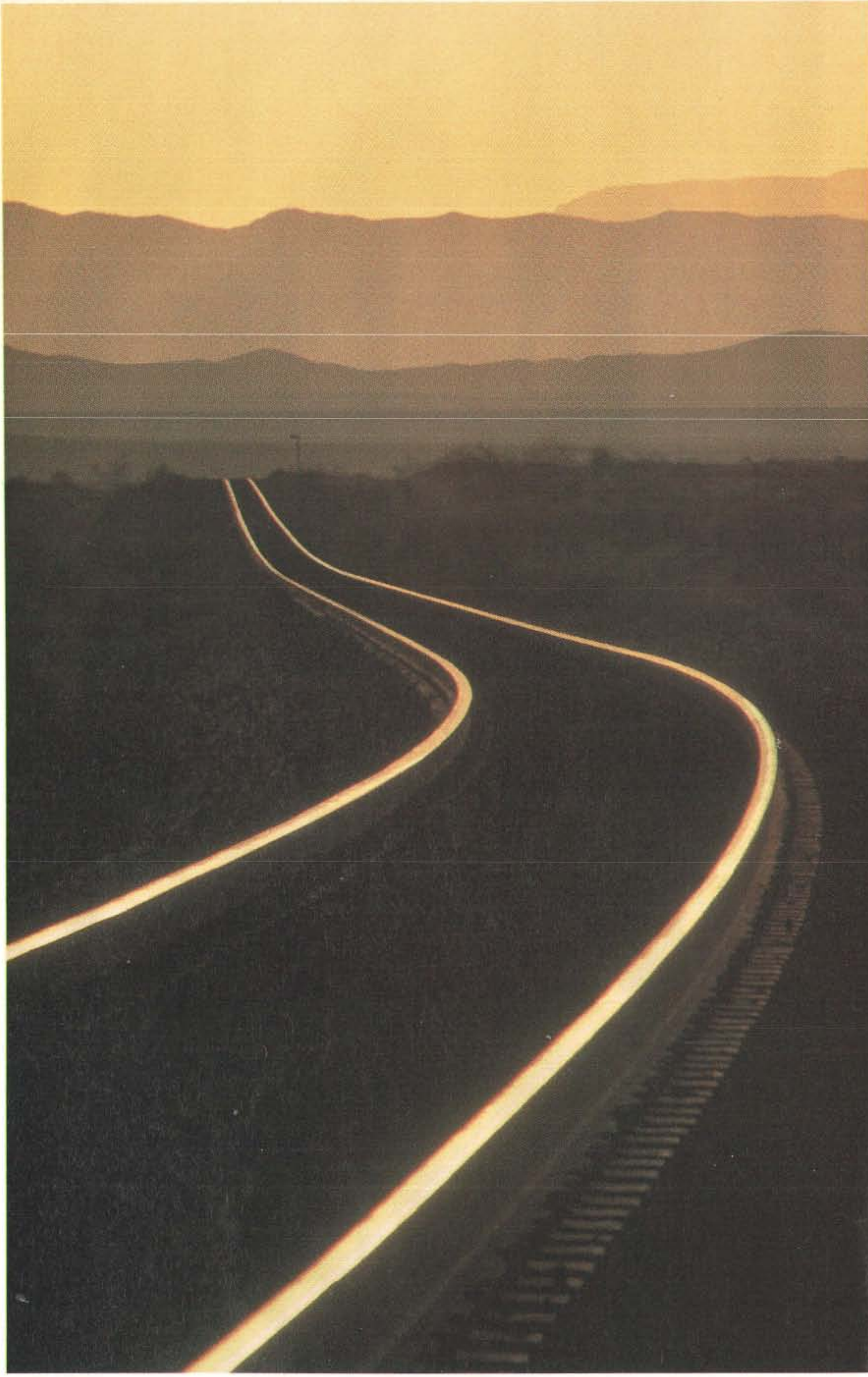
FURTHER READING

SECURITY WITHOUT IDENTIFICATION: TRANSACTION SYSTEMS TO MAKE BIG BROTHER OBSOLETE. David Chaum in *Communications of the ACM*, Vol. 28, No. 10, pages 1030–1044; October 1985.

THE DINING CRYPTOGRAPHERS PROBLEM: UNCONDITIONAL SENDER AND RECIPIENT UNTRACEABILITY. David Chaum in *Journal of Cryptology*, Vol. 1, No. 1, pages 65–75; 1988.

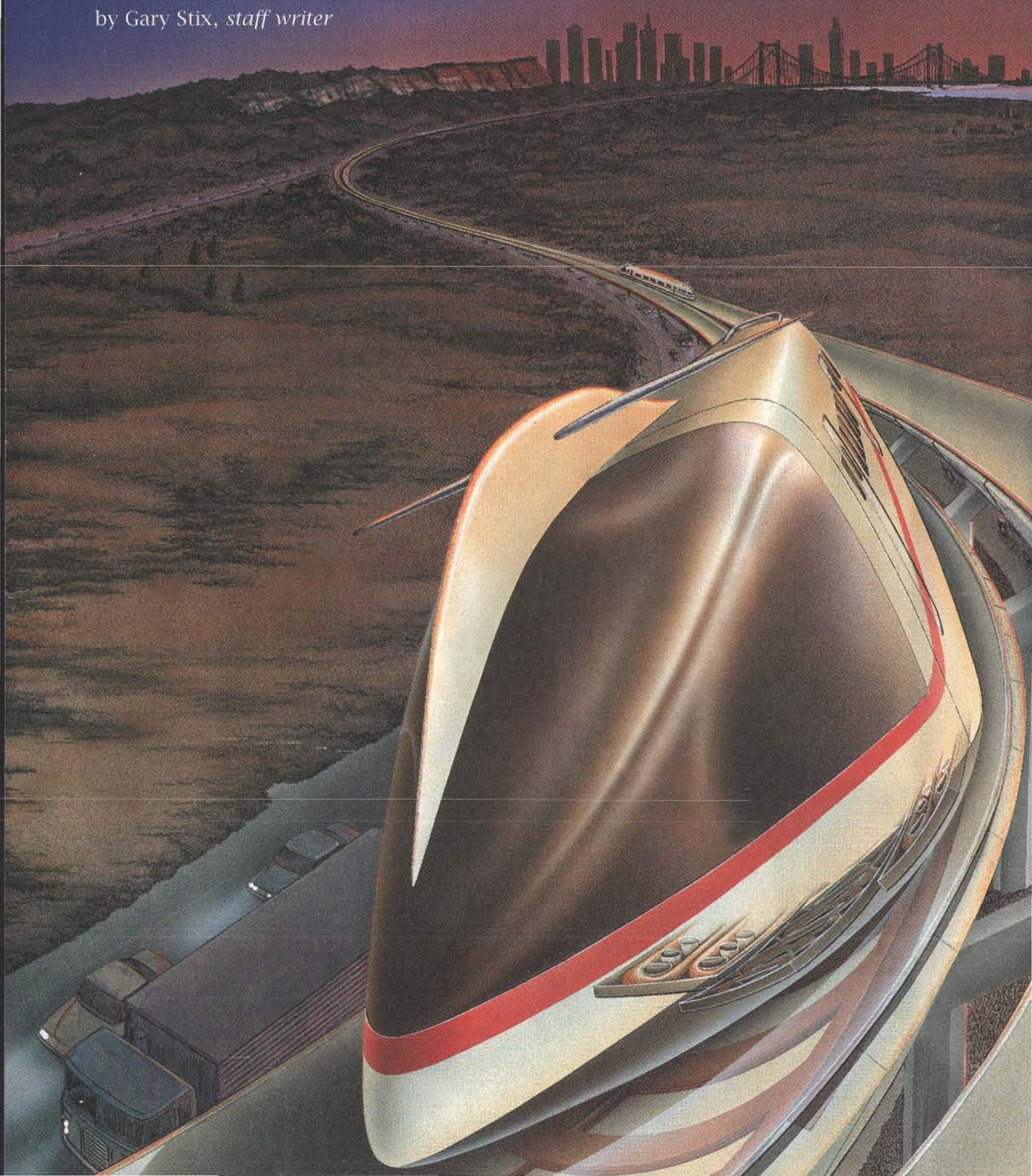
MODERN CRYPTOLOGY: A TUTORIAL. Gilles Brassard in *Lecture Notes in Computer Science*, Vol. 325. Springer-Verlag, 1988.

PRIVACY PROTECTED PAYMENTS: UNCONDITIONAL PAYER AND/OR PAYEE UNTRACEABILITY. David Chaum in *Smart Card 2000: The Future of IC Cards*. Edited by David Chaum and Ingrid Schaumüller-Bichl. North-Holland, 1989.



AIR TRAINS

by Gary Stix, *staff writer*



The nearly 100-year-old dream of building a vehicle able to fly atop a magnetic cushion has captured the imagination of a U.S. senator and a few graying engineers who predate the pocket calculator.

Henry H. Kolm hates trains. He bristles when someone has the indelicacy to talk about his magnetic flying machine in the same breath as a train. It does have a passenger cabin and travels on—or rather over—a track of sorts. Just don't call it a train. Trains, in Kolm's humble view, were the fundamental breakthrough that moved transportation a step beyond the stagecoach. "Tremendous technology for the year 1837," says the former Massachusetts Institute of Technology research scientist whose signature article of clothing is a bolo tie.

With an accent that lingers from a boyhood in Austria, the 67-year-old Kolm carefully explains that his invention, magneplane, is a creature of this century or perhaps the next. It is the idea of someone, like himself, who has piloted airplanes commercially and who has discarded Victorian railroad ties and ballast for concepts such as "coordinated banking" that only an aviator would know.

Magneplane is more than a run-of-the-mill aircraft, too. It uses magnetism, not aerodynamics, to lift a 50-ton vehicle to a maximum altitude of six inches off the ground. Once aloft, passengers can then race at the speed of a fast turboprop aircraft down a troughlike guideway that looks more like a bobsled run than a train track. An intercity network of such vehicles capable of traveling at speeds in excess of 250 miles an hour could compete with air travel. For Kolm, high-speed rail, such as France's Train à Grande Vitesse, would pale by comparison.

More from the likes of Buck Rogers than from Casey Jones, Kolm's notion of a magnetically levitated vehicle, or more simply maglev, was the most radical of a number of U.S. designs that date from the 1960s—and that have been gathering dust in university and corporate archives until the technology staged a recent revival.

The vision of Kolm and others has enraptured a pugnacious U.S. senator who is partial to bow ties and tweed Irish walking hats. Last year, Senator Daniel Patrick Moynihan of New York ramrodded a \$725-million provision for maglev development into the mammoth, six-year, \$151-billion highway bill.

The maglev portion of the bill is a nose-thumbing tweak to the administration of George Bush, which has yet to shake a deep-seated aversion to any form of government meddling that encourages development of one commercial technology over another. The language for maglev in the highway bill begins with the blunt assertion: "It is the policy of the United States to establish in the shortest time practicable a United States designed and constructed magnetic levitation transportation technology."

RUSHING PAST RUSH HOUR, magnetically levitated trains could move passengers at speeds five times that of the highways below them if the trains can be tamed into something more than a flying roller coaster.

HENRY H. KOLM, one of the original U.S. maglev developers, leans on his Piper Navajo Chieftan, which he flies to technical conferences instead of taking the train.

Ever the New York City street kid cum Harvard intellectual, Moynihan framed the maglev program as an underdog competition with Germany and Japan. Although the U.S. did pioneering work in maglev, the domestic effort ground to an abrupt halt in 1975, when the White House Office of Management and Budget (OMB) decided that money for the technology was better spent elsewhere. "That was the accounting office making the decision," Kolm still rasps after nearly 20 years. Meanwhile the U.S.'s two main economic rivals have each sunk about $1 billion into maglev research since 1969. A German consortium is even ready to build a maglev line from the Orlando, Fla., airport almost to the door of Walt Disney World.

Sizable maglev funding in the highway bill, which is intended to draw an additional $200 million in investment from private industry, would put the U.S. back in the maglev game. Presumably, the quirky inventive ideas of a Henry Kolm could mix with the dormant design and manufacturing expertise of the financially strapped U.S. aerospace and civil engineering companies that thrived when building jet fighters and erecting nuclear power plants were national priorities. The term "maglev," coined 25 years ago at the height of the Vietnam War, was immediately adapted to the spirit of those times. "Mag Lev Not War," declared one engineer.

While President Bush signed the highway bill last December, his administration had by no means issued its final pronouncement on the matter of spending this much on a new transportation technology. The administration, including Kolm's bane, the OMB, has requested no appropriation for maglev development for the 1993 budget year, which begins in October. At the very least, studies of prospective ridership, financing and technical feasibility now being carried out by several federal agencies should be completed before the program advances toward building a prototype. Maglev funding appropriations also face opposition from some in Congress who view the technology as too outlandish.

On the sidelines a number of sober-minded transportation professionals who are watching the effort to advance maglev are experiencing a distinct sense of déjà vu. The past quarter century is littered with failed plans that promised to revive the nation's anemic passenger rail service. As pursestrings are pulled ever tighter, some transportation specialists ask whether this type of swashbuckling foray into accelerated technology development could turn into a high-tech boondoggle. Maglev is a funny kind of plane. But it is one that the Germans and the Japanese have invested in for more than 20 years, and their work has not yielded a fare-paying passenger. Ridership is uncertain. And a maglev program could divert money away from the more practical prospects of upgrading intercity rail service.

To Moynihan, maglev makes perfect sense. As chairman of a Senate subcommittee with jurisdiction over federal highways, he has advocated maglev to help move the U.S. beyond decades of frenzied, sometimes uncontrolled, road building. In 1960 he wrote that the National System of Interstate and Defense Highways, better known as the interstates, was "throwing up a Chinese wall across Wilmington, driving educational institutions out of downtown Louisville, plowing through the center of Reno. When the interstate runs into a place like Newburgh, New York, the wreckage is something to see."

The highway bill, which no longer bears the word highway in its name (also Moynihan's doing), provides $7 billion for the completion of the 43,000-mile crisscrossed interstate network, primarily for stretches of road in Boston and Los Angeles. Moynihan noted that the interstates had exacerbated the traffic congestion they were designed to relieve—and the problem will only grow worse. "By 2020, it will take 44 lanes to carry the traffic on I-95 from Miami to Fort Lauderdale," Moynihan wrote to voters. "Pretty soon there won't be anything left of Florida!"

Something was needed, he reasoned, to move beyond this national obsession. "We have poured enough concrete," he proclaimed. One answer was to make use of the epic legacy of the Eisenhower administration: put narrow elevated tracks alongside the pavement of an interstate to move as many people as can drive on eight to 10 lanes. Traffic on these guideways need not respect the speed limit below. The more-than-300-mile-per-hour top speed attained by a maglev might make New York's Metropolitan Opera into an evening sojourn for the well-to-do who live by the Potomac.

Moynihan began his dogged pursuit of a domestic maglev program about five years ago, when he wrote a maglev bill that went nowhere. He then impaneled a study group of enthusiasts, many of whom had experienced the frustrations of the earlier program. Kolm, of course, was a member. The panel's not surprising conclusion: "Maglev is a revolutionary concept that will one day—sooner than we think—be a reality."

In 1989 the panel—the Maglev Technology Advisory Committee—predicted that maglev would reduce highway and airport congestion, consume half the energy per passenger mile burned by the automobile and a quarter of that used by the airplane. The technology would lessen dependence on imported oil (maglev can use electricity generated from coal, nuclear or hydroelectric sources) and diminish emissions of hydrocarbons and other pollutants.

Moynihan used the report as a prod to a somnolent federal bureaucracy. He sought $1 million for the U.S. Army

Corps of Engineers to prepare a plan for a domestic maglev program. The Department of Transportation (DOT), whose response to Moynihan's previous inquiries had been lukewarm at best, suddenly climbed on board. The DOT undertook a feasibility study, called the National Maglev Initiative (NMI), a joint project that combined its own expertise with that of the Corps of Engineers and the Department of Energy.

Moynihan did not wait for the expected equivocal response to this multiagency endeavor, due out early next year. Instead he inserted the $725 million into the once-in-six-year cornucopia of the highway bill. In fact, the National Magnetic Levitation Prototype Development Program received more money in the erstwhile highway bill than did an advanced technology research program for highways. "This kind of money has never been seen before for maglev in the United States," marvels John T. Harding, an official in the DOT's Federal Railroad Administration and the chief scientist for the NMI.

Magnetic Magic

The idea of riding atop a surflike magnetic wave is almost as old as this century. A scheme for magnetically levitated transportation was proposed in November 1909 in the pages of this magazine. Robert Goddard, the rocket pioneer, suggested that a tunnel could be built from Boston to New York in which cars would ride in a partial vacuum while suspended and propelled by the "magic power of magnetism."

It was only a few years later, in 1912, that Emile Bachelet, an expatriate French engineer, patented the first design that somewhat resembled current maglev prototypes. Bachelet's 33-pound model, which once crashed through the wall of his laboratory in Mount Vernon, N.Y., levitated above a continuous row of electromagnets supplied with alternating current. But it consumed too much power to ever become a practical method for lifting a 40-ton vehicle.

Then, almost 50 years later, James R. Powell, an engineer from Brookhaven National Laboratory who designs nuclear reactors, came up with a workable solution. Powell was on a weekend jaunt to visit his girlfriend in Boston when he ran into a monstrous traffic jam on the approach to the Throgs Neck Bridge. "It took four hours to get over the bridge, a total distance of three miles," Powell says. "I kept thinking there must be a better way to travel long distances. It was then that I thought of maglev."

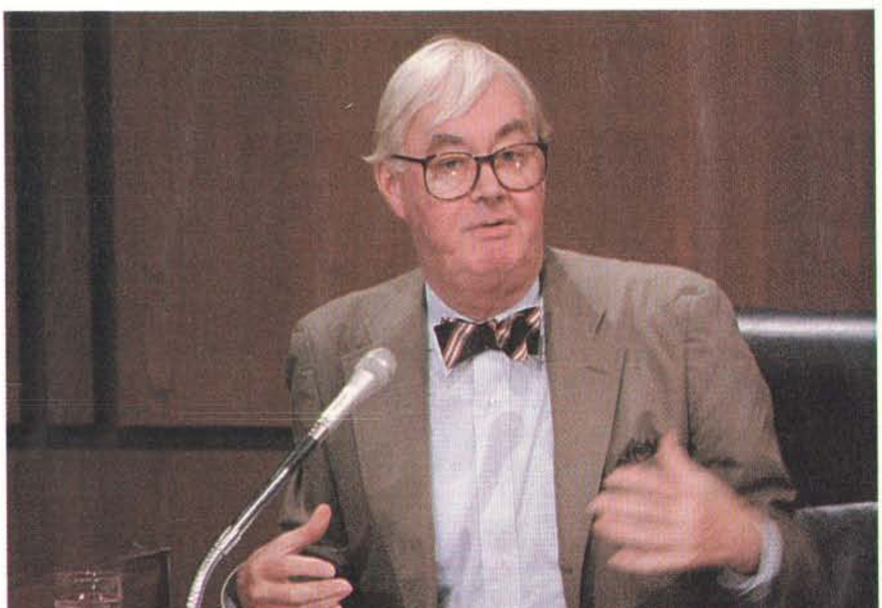
Powell, who was unacquainted with Bachelet's work, also borrowed the grade school principle that magnets of like polarity repel. Refining his idea with Brookhaven physicist Gordon T. Danby, he proposed using a magnet on the vehicle to generate current in metal coils set into the guideway. Because this current produces a magnetic field with the same polarity of the vehicle's magnet, the two fields repel each other. Powell and Danby's main insight was that sufficient lifting force could be obtained only by using superconducting magnets, the kind Brookhaven employs to focus the beams of particle accelerators. The scheme called for coils of superconducting wire on the side or bottom of the vehicle that were cooled to four kelvins with liquid helium. The magnetic fields were strong enough to lift 40 to 50 tons. The interaction between the magnets and the coils also serves to stabilize the vehicle.

Once suspended, this air train still has to move forward and backward. It does so with a linear synchronous motor. The guideway coils are equivalent to the ringlike stationary electrodes, or stators, in a conventional rotary electric motor. The same magnets used for levitation are pulled, then pushed by fields of alternating polarity generated by current supplied to magnetic coils in the guideway. Speed and acceleration are controlled by varying the frequency and the intensity of the current. The two researchers presented their work in 1966 in a paper entitled "High-Speed Transport by Magnetically Suspended Trains."

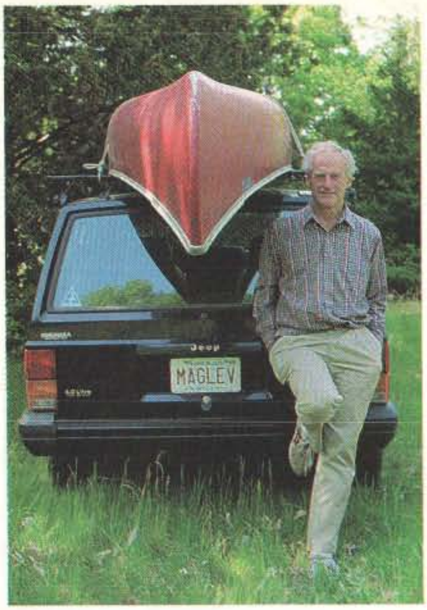
In the mid-1960s, just as now, some elements within the political establishment came to view advanced technology as a panacea for reviving moribund intercity passenger rail service, and Congress passed the High-Speed Ground Transportation Act of 1965. The program first focused on a terrestrial version of a hovercraft that moved along tracks.

As the work of Powell, Danby and other researchers became known, interest shifted to maglev. The Stanford Research Institute built a half-ton test sled that levitated with superconducting magnets. Devised with Richard D. Thornton of M.I.T., Kolm's magneplane was perhaps the most novel proposal. With help from Raytheon Company, 80 graduate students and a $650,000 grant from the National Science Foundation, Kolm and Thornton built a 1/25th scale model that raced down the trough guideway beginning in 1973. Those efforts and others arguably gave the U.S. the lead in maglev research. Then the OMB pulled the plug, and Japan and Germany forged ahead.

In 1970 Japanese National Railways appropriated a design almost identical to that conceived by Powell and Danby. Over the years the Japanese program has produced a series of vehicles that



SENATOR DANIEL PATRICK MOYNIHAN of New York has imagined maglevs racing above the interstates, which he views as a checkered legacy of decades of frenzied road building that tore apart urban neighborhoods.



RICHARD D. THORNTON (*left photograph*) collaborated with Henry H. Kolm during the mid-1970s. Thornton is now pursuing ideas for a U.S. maglev design on his own. Gordon T. Danby and James R. Powell, researchers at Brookhaven National Laboratory (*left to right in right photograph*), published the first paper on a maglev using superconducting magnets in 1966.

have been tested on a 4.4-mile experimental track in Miyazaki Prefecture in the southwest of Japan. The Railway Technical Research Institute, which is funded by the now privatized railroad, has been planning a nearly 27-mile test line in Yamanashi Prefecture at a cost of about $2 billion, which could be extended within 20 years' time into a route connecting Tokyo to Osaka.

The Japanese vehicles may need at least five more years before they are ready to carry commercial passengers. The maglev that Japanese engineers envision would run at 310 miles per hour and could carry 10,000 people per hour in trains up to 14 cars long. The Tokyo to Osaka run, already the nation's most heavily traveled train route, would require only an hour, a third of the time taken by the bullet trains that began service in 1964. But if development lags, a new bullet train line could replace the maglev route.

The still experimental nature of the Japanese program was underscored last October, when the MLU002, the most advanced prototype vehicle, was consumed in a fire. Institute officials say the fire that ignited in a magnesium-alloy wheel component does not undermine their basic approach to magnetic levitation and propulsion. If anything, it proved the durability of the superconducting magnets. An hour after the blaze, a firefighter approached the vehicle, wielding a fire ax. The metal implement suddenly flew from his hand toward a still active magnet.

The only maglev that can exceed 200 miles per hour and is close to installing a fare box is of German pedigree. But Germany's Transrapid 07, the seventh prototype in a program that began in 1969, also faces obstacles. After testing several approaches, Germany's transport ministry decided in 1977 on a vehicle based on a design first devised during the 1920s by German scientist Hermann Kemper. Instead of levitating above a track, the bottom of the train wraps around the top of a T-shaped guideway. A conventional electromagnet is placed on the inside section of the vehicle that curls underneath the guideway. The underslung magnet is attracted upward to within less than an inch of a steel rail in the guideway. As with the Japanese maglev, the vehicle is propelled by a linear synchronous motor.

The billion dollars or so of government money sunk into the project thus far has been supplemented by private investments from Transrapid International, a consortium of three German companies. Although the German government has granted Transrapid partial technical and safety certification, other tests remain. No one has done anything more than simulations of the possible destabilizing aerodynamic effects of two vehicles passing each other in opposite directions at 250 miles per hour, separated by only a few feet.

Nor has the German public put out a financial welcome mat for the Transrapid. The state of North Rhine-Westphalia rejected using even a "pfennig" of government funds for a route that would have connected the Cologne/Bonn and Düsseldorf airports and the city of Essen, effectively blocking the project. Environmentalists also objected to the prospect of guideways towering over the landscape and noise from the vehicles.

What has kept Transrapid alive at home is reunification of the two Germanys. Like Moynihan, Günther Krause, the transportation minister who hails from the former East Germany, sees maglev as a gleaming emblem of renewal, a physical linkage that symbolizes future parity between the former East and West. It is Krause's dream to spend $5 billion to make maglev into a 21st-century version of the "Flying Hamburger" that connected Hamburg and Berlin in the 1930s. *Der Fliegender Hamburger* reached speeds of a little more than 100 miles per hour, faster than the trains that make the trip today. Private financing must still be procured to supplement the government's contribution.

While Germany seeks backing for this public-private venture, the first commercial line is likely to be built in the U.S., which is seen as perhaps the primary market for the Transrapid. The nearly 14-mile route between the Orlando airport and a station near theme parks such as Walt Disney World is scheduled to open by 1996. Senator Bob Graham of Florida wrote nearly $98 million into the highway bill for this $600-million project, a collaboration between Japanese, German and U.S. companies. Amtrak would operate the line.

The Florida maglev will be a closely watched train, an early indicator of whether the technology will prove financially viable anywhere. "People visiting Florida will be on vacation, interested in

a novelty ride and willing to pay slightly more," says Edith B. Page, a former senior associate with the congressional Office of Technology Assessment (OTA) who supervised an agency report on maglev. "If maglev doesn't make it in Orlando with existing technology, it's going to be very difficult to find the right niche for it."

The U.S. is counting on the interagency National Maglev Initiative to weigh the feasibility of designing and producing a domestic maglev or, alternatively, licensing the technology from abroad. Reports from the U.S. Army Corps of Engineers, the Department of Transportation and the Department of Energy as well as a number of private contractors are due to be published by the spring of 1993. Analyses range from revenue and ridership prospects to the impact of noise and electromagnetic fields. Four NMI teams are trying to sketch out ideas of what a uniquely American maglev would look like and how U.S. technology could surpass the prototypes already assembled by the Japanese and Germans.

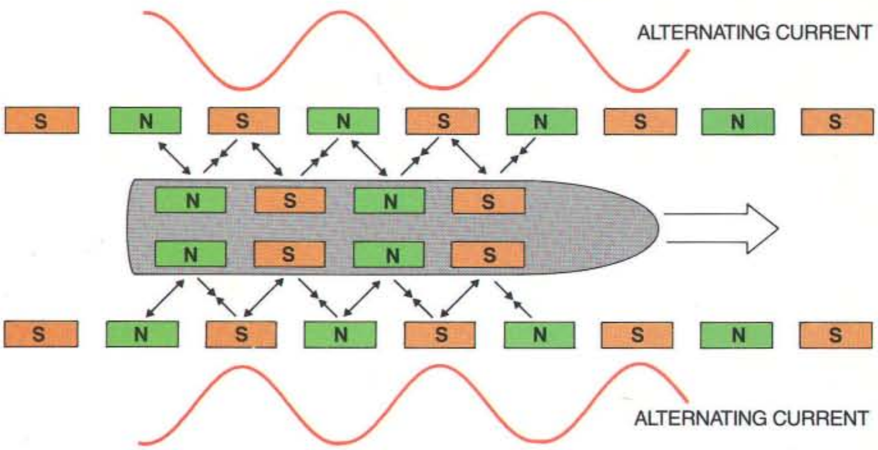
One team is headed by Grumman Corporation from Moynihan's home state of New York. Grumman worked on a 1960s-vintage hovercraft and has gathered six other engineering organizations onto the team. Even before the results of the NMI are released, Grumman has joined a group of companies that is attempting to put together a plan for a possible maglev line from Washington, D.C., to Baltimore. Richard J. Gran, who launched the company's maglev effort, helped to design the digital flight control system for the capsule that descended to the moon during the Apollo program. "That was several orders of magnitude more difficult as a technical feat than building a maglev," he comments.

The Grumman maglev proposal for the NMI is a hybrid of the German and Japanese work. Like the German design, the vehicle wraps around the guideway, preventing it from leaving the track. The vehicle levitates at all speeds and so does not need wheels for takeoff, as does the Japanese prototype. And the level of magnetic fields produced by the Japanese maglev, enough to affect a pacemaker in some areas within the vehicle, is minimized in both the German and Grumman vehicles. The fields in the German and Grumman systems are absorbed by the iron in the vehicle magnets and by the steel in the guideway. The possible health effects of electromagnetic fields could become an issue in operating maglevs.

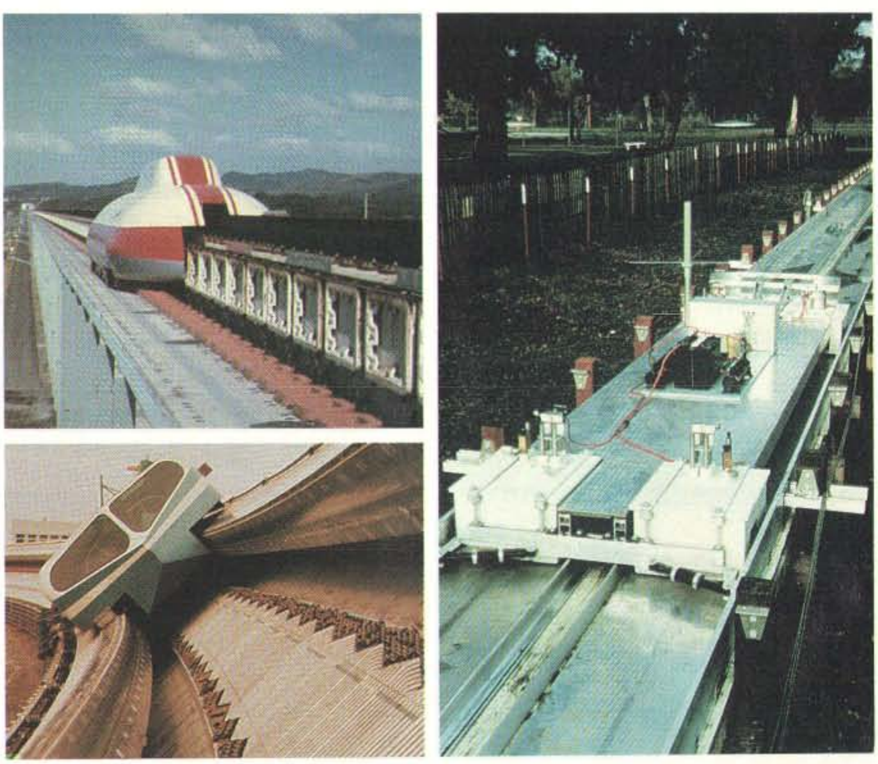
Taking these design elements as a starting point, Grumman has tried to make improvements on the basic German system. Levitation using attracting magnets is inherently unstable. Once activated, the magnets that wrap around the bottom of the guideway are pulled upward toward a steel rail. If the magnets get too close, they will adhere to the rail; if they drift too far away, they will be pulled down by gravity.

In the Transrapid system the magnets are only powerful enough to maintain an air gap of about three eighths of an inch between the train and the guideway. Each stretch of guideway must be made almost as smooth as glass so that the vehicle does not graze the metal, an obvious safety concern at high speed. These precision tolerances could be misaligned by wind, temperature variations or vibration, which might require frequent and costly adjustment.

Instead Grumman proposes to avoid the tight clearances by combining small, conventional magnets with superconducting magnets similar to those employed by the Japanese. An electric current in the nonsuperconducting magnets can be varied by a digital control



LINEAR SYNCHRONOUS MOTOR employs alternating current to generate a magnetic wave that travels along with the vehicle. These time-varying fields interact with the vehicle magnets to push and pull the vehicle along. This design is from a Japanese prototype. Most other maglevs are fitted with their own version of these motors.



1970-VINTAGE MAGLEV PROTOTYPES are shown from Japan (*upper left*), from Germany (*bottom left*) and from the Stanford Research Institute (*right*) in the U.S.

system to preserve the separation between the vehicle and the guideway. The high overall magnetic flux produced by the superconducting magnets allows a separation of two inches to be maintained, avoiding the need for frequent realignments of the guideway. "These are bridge-type tolerances," Gran says.

A larger separation between the vehicle and the guideway is a design goal of other NMI developers as well. But Transrapid officials assert that Grumman and other prospective developers may discover, as did the Germans, that a small gap between guideway and vehicle is needed to ensure a smooth ride. "The Transrapid concept is the right concept; most other ideas were discarded 10 or 20 years ago in Europe," claims Manfred Wackers, president of Transrapid International.

Make or Buy

Because up to 80 percent of the costs of a maglev system may go to build the guideway, all the NMI teams are trying to prove that they can reduce these multibillion-dollar capital investments by designing lightweight, inexpensive and easy-to-assemble structural supports. Bechtel, which has been involved with civil engineering projects ranging from railroads to interstates to nuclear power plants, leads an NMI team that has proposed using a guideway consisting of a narrow, hollow box beam about five feet wide.

Minimizing materials and labor could bring costs down to between $10 million and $15 million a mile, half of some U.S. developers' estimate for the Transrapid guideway and less than the expense for a new mile of highway in some urban areas. Wackers of Transrapid asserts that the cost for a two-track guideway would range from $20 to $25 million. U.S. proposals, he asserts, may prove even more expensive. "Nobody in the U.S. has experience manufacturing guideways; they're just drawing pictures of cross sections."

Not everyone sees the necessity for an all-American maglev. Richard A. Uher, a Carnegie Mellon University researcher who worked for Westinghouse Electric when the company was involved in studies of high-speed trains in the late 1960s, believes that the desire to better the German and Japanese work is misguided. Uher contends that the NMI will turn up no more than cosmetic improvements to the more than 20 years of unbroken overseas research.

During the mid-1980s, Uher helped to organize a group of steel producers, unions and community groups to form Maglev, Inc., a corporation that would use technology licensed from Transrapid to create a regional maglev manufacturing industry around Pittsburgh. Uher's group wants to use the money authorized in the highway bill to build a 19-mile route from the Pittsburgh airport to the city center, the first link in a network that would gradually spread from Pittsburgh to eastern Pennsylvania, Ohio, West Virginia and Maryland. "It doesn't matter where a technology is invented; it matters where it is manufactured," Uher says.

Whether it is homegrown or not, riding a high-speed maglev will be a very different experience from taking a Sunday drive in the country. The maglev provision in the recent highway bill would use the land adjoining the interstates to build the looming guideway structures. But curves on the interstates were created for automobiles traveling at maximum speeds of 70 miles an hour.

A train traveling three to four times that speed must either slow down and speed up frequently or diverge a mile or so from the interstate to avoid subjecting a passenger to unreasonable gravity forces. "You can't track the centerline of the highway within comfortable parameters without making the ride like a roller coaster," says Steven G. Carlton, who heads a study at Martin Marietta for the NMI.

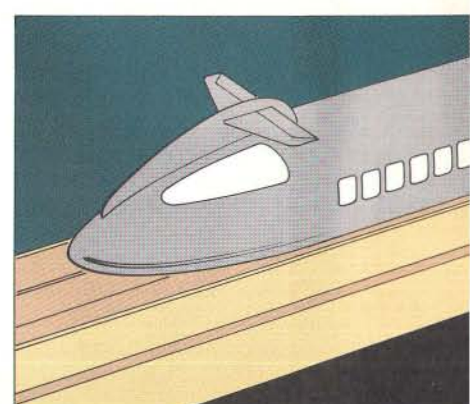
Other experts argue that passenger discomfort caused by the effects of centrifugal forces can be offset by banking the track or tilting the train. Grumman has proposed banking the guideway up to 15 degrees and allowing the vehicle to tilt another nine degrees. All but four of the turns on the New York State Thruway could then be negotiated at 250 miles per hour without leaving the highway area.

Kolm's magneplane [see "Electromagnetic Flight," by Henry H. Kolm and Richard D. Thornton; SCIENTIFIC AMERICAN, October 1973] forwards the most extreme notion. Magneplane's curving trough over which the vehicle glides like a bobsled allows coordinated banking at angles of 45 degrees—equivalent to an aircraft maneuver of simultaneously adjusting the rudder and aileron to turn in the vertical (yaw) and longitudinal (roll) axes. Kolm suggests that the experience may be one that people will grow into. "If Mozart woke up and rode in an automobile, he might find it disconcerting. But after 10 times he might find it enjoyable," Kolm says.

Researchers at the Department of Transportation's John A. Volpe National Transportation Systems Center in Cambridge, Mass., are taking passen-

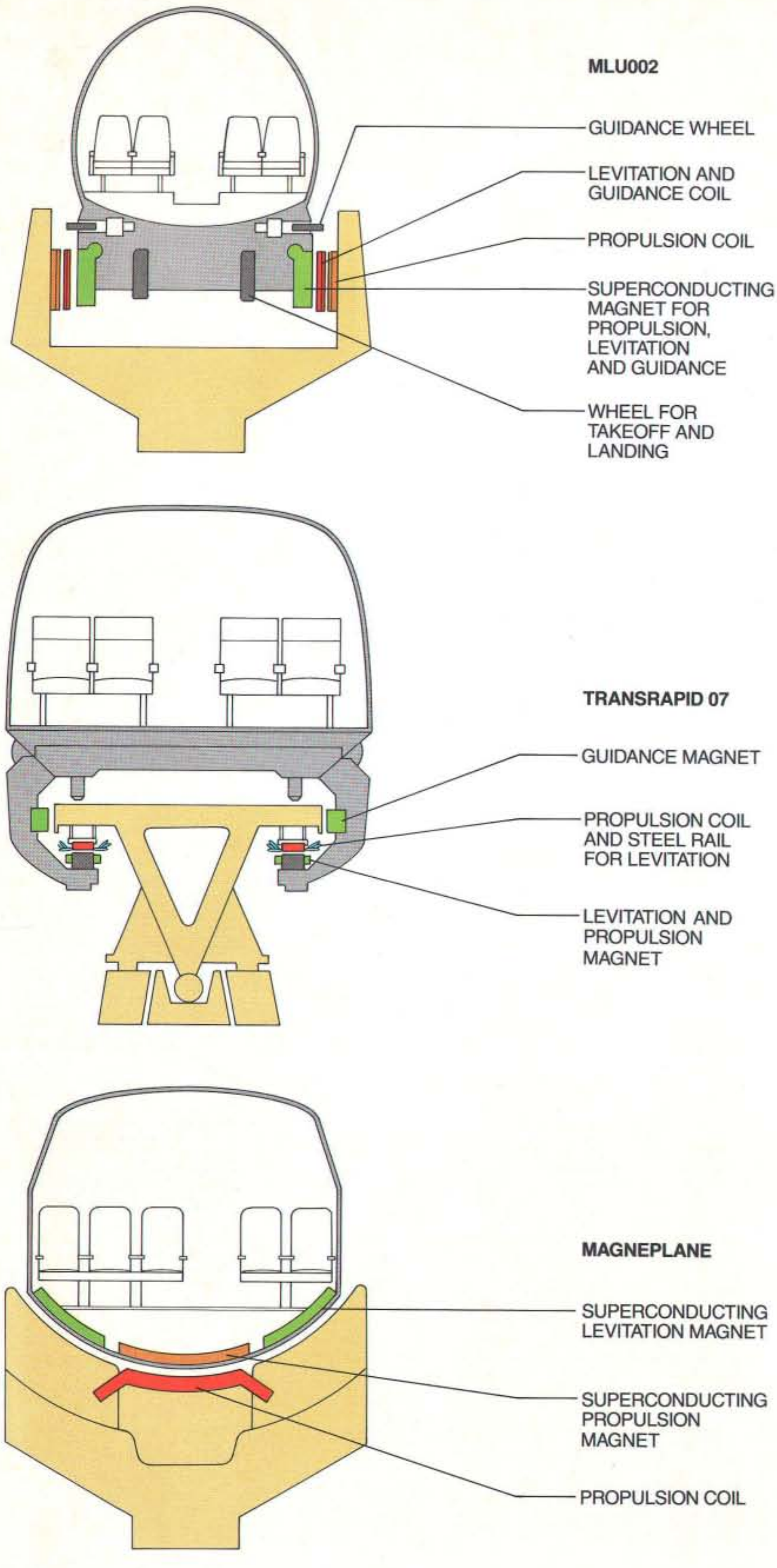
MLU002, a Japanese maglev prototype destroyed in a fire last year, has superconducting magnets that interact with metal coils in the guideway's side walls. Magnetic fields induced in the coils create a repelling force that lifts and stabilizes the vehicle four inches above the guideway.

TRANSRAPID 07, the German maglev that is being readied for commercial service, is built with nonsuperconducting magnets attached to the vehicle. These magnets are attracted upward to within an inch of a steel rail on the guideway.



MAGNEPLANE'S superconducting magnets induce repelling forces in the aluminum-sheet guideway, which levitates the vehicle by six inches. The trough-shaped guideway allows banked turns of up to 45 degrees.

Exercises in Levitation



gers in a leased corporate jet to expose them to the type of ride that would be experienced when careering back and forth into 25-degree banked turns. "To be blunt about it, these systems will not provide the ride of a 747 at 40,000 feet," says E. Donald Sussman, chief of the operator performance and safety analysis division at the Volpe center. "The question is what is needed for passengers not to reject it."

Another concern is the noise from a maglev, particularly the "startle effect" of a train traveling at 300 miles per hour suddenly bearing down on motorists or a homeowner raking leaves. At top speed, noise reaches a deafening 100 decibels 80 feet from the guideway, substantially higher than the roar of a heavy truck passing a motorist driving with the car windows down, according to Carl E. Hanson, a consultant for the NMI. A maglev nearing a city at speeds below 200 miles per hour registers a more tolerable 80 decibels, about the equivalent of a passing city bus.

Deciding on what G forces an infirm couple can withstand when taking the weekend maglev shuttle to visit the grandchildren may be a simple proposition compared with the institutional arm twisting that will be needed to rally the government behind a radically new means of transportation. Passenger railroad travel in the U.S. is negligible, accounting for 1 percent of trips between cities. As a consequence, the Federal Railroad Administration in the Department of Transportation has had little experience in administering ambitious technology research and development programs for passenger trains. In fact, much of the agency's time is spent enforcing safety regulations.

So Moynihan's subcommittee, which has jurisdiction over the U.S. Army Corps of Engineers, set out to enlist the agency's civil engineering expertise. In an address to the Senate last year, Moynihan referred to the corps as a valuable but sometimes poorly utilized resource. Why, he asked, had the corps designed and supervised more construction in Saudi Arabia than in the U.S. during the early 1980s? "In the course of this foreign adventure, we built, among other things, 46 mosques," he said. "That's excellent as a limited exercise, but even mosque building can be carried to excess."

Moynihan's vision is not necessarily in step with that of the administration. Maglev occupies one page in the national transportation policy, a 129-page document whose bright photographs give it a marked resemblance to a corporate annual report. The document, published by the Department of Trans-

portation in 1990, embraces the concept of intermodalism—planners' jargon for getting different forms of transportation to complement one another—a train-to-the-plane, for example.

If one looks more closely, the administration's readiness to move beyond the decades-old reliance on automobile and air transport is equivocal at best. The policy restates the opposition to paying for the shortfall in Amtrak operating costs that are not met by passenger fares and for capital improvements for 40-year-old rail cars, a hard-edge position that has now begun to ease slightly. The policy document further emphasizes that private investment will be the "central feature" in building a maglev.

Transportation analysts question whether the huge capital outlay for any high-speed rail project can be economically justified under present policies. In a report issued last year, the congressional Office of Technology Assessment compared maglev with other means of intercity transportation, including jets, conventional trains, other forms of high-speed trains and tilt-rotor aircraft (those that take off like a helicopter and fly like an airplane). On a New York to Washington, D.C., route, maglev registered the highest capital costs, some $7.4 billion for both vehicle and the guideway using the Transrapid. Over time, though, maglev would have the lowest operating expenses of the various technologies surveyed. The 3.4 cents per mile for each train is about half the cost per seat-mile for a jetliner.

Even so, Edith Page, former OTA official, told Congress that spending a billion dollars on maglev technology would fail to yield a U.S.-developed system that would be ready to carry passengers at the end of six years. Even if a maglev route were ready, people might not use it. "People make travel decisions

A Train without Seat Belts

Maglev is an endorsement of a technology with much promise and little proof. Not so other forms of high-speed rail transportation. Japanese bullet trains, the French Train à Grande Vitesse (TGV) and the like have, in total, carried billions of passengers at top speeds approaching 200 miles per hour—virtually without mishap. Europe is planning a continent-wide network of high-speed rail links, extending all the way to eastern Europe.

Unlike a maglev, the TGV would be familiar to the railroad builders of the early part of the 19th century. A TGV could pull into Union or Penn Station on standard-gauge tracks. It also retains the character of an old-fashioned train, replete with dining car and lounges. A seat belt is nowhere to be seen.

The U.S. may soon get to judge the merits of conventional high-speed rail against maglev. A maglev project in Orlando, Fla., is scheduled to begin operation in 1996. Meanwhile the state of Texas last year granted a franchise for a high-speed rail route to a consortium led by Morrison Knudsen, the civil engineering and construction firm. The group, which includes GEC Alsthom, the manufacturer of the TGV, plans to build high-speed links between five Texas cities by the year 2000. One small detail: almost all of the $7 billion in private financing must still be arranged.

To some, these conventional rail systems hold a technological edge over maglev. By attaining a speed of 320 miles per hour in 1990, the TGV nearly achieved the world speed record for a train, faster than any maglev, except for an unmanned Japanese prototype. Although the record was set on a test run, the French National Railways wants to lay down sections of new TGV track to handle trains running at 250 miles per hour, about the speed projected for many maglevs. "At the moment, there is nothing that maglev can do that rail can't do better," asserts Vukan R. Vuchic, a professor of transportation engineering at the University of Pennsylvania. Whether higher speeds are really needed is an open question. As speed mounts, say, from 200 to 300 miles per hour, time savings dwindle, and energy costs rapidly increase.

TGV reaches 320 miles per hour.

The maglev half of this debate demands a leap of technological faith. Although the conventional train has traveled faster than anyone could ever have foreseen, maglev enthusiasts still contend that this technology is nearing the limits of about 175 years of development: ramming a 300-ton train down an unbending stretch of downhill track at more than 300 miles per hour will overstress both track and train in actual day-to-day service. "I wouldn't have liked to have been a passenger on the TGV that made the record," says Manfred Wackers, president of Transrapid International, the maglev developer. "You should have seen the film of how the wheels were dancing on the rail and the sparks coming from the catenary. This is not something you can have in daily operation."

Technical challenges abound in making still roughshod maglev designs into a technology that passengers will pay to ride. In some ways, it is like molding a roller coaster into a public transportation vehicle. But Henry H. Kolm and other maglev designers contend that, even if the speed factor is discounted, maglev displays qualities unknown to the railroad designer: rapid acceleration and braking, an ability to climb steep grades and bank at sharp angles on narrow, curving guideways, operation with single vehicles at tight time intervals like an automobile on the freeway, and low maintenance costs because of lack of contact between the vehicle and the guideway.

If, in the end, the technology does not fly commercially, a turn-of-the-century theme park may buy the patents.

based on cost and convenience," Page says. "It still isn't clear that this technology is a cost-effective competitor to the automobile on the low end, the train in the mid end and air on the high end." (Page recently left the OTA to work in Bechtel's Washington office.)

Whether a maglev—or any type of high-speed train—would attract enough riders is another uncertainty. A committee of the National Research Council tried to tackle the question of whether a high-speed rail system could pay its own way under a highly optimistic assumption. The study projected what would happen if all passengers who travel the most heavily used air routes switched to a high-speed train, while still paying airline fares. Its conclusion: revenues from only the most traveled of the routes studied—San Francisco to Los Angeles—would be enough to exceed the capital and operating expenses.

Maglev will not become a reality without a federal helping hand. Land grants helped to build the railroads. Trust funds raised billions of dollars for airport and highway construction out of ticket and gas taxes.

A blueprint for financing new transportation technology may be found in the highway bill. The bill, much of which bears the imprint of Moynihan, has now taken on the ponderous name of the Intermodal Surface Transportation Efficiency Act of 1991, sometimes contracted simply to Ice Tea in speech. A reader has to move past the cover to find the word "highway." The bill takes the unprecedented step of allowing state and local planners to channel a substantial percentage of trust money set aside for roads to other forms of transportation. Under the new law, up to nine out of every 14 cents of federal tax paid on a gallon of gas can be diverted from highway programs, for everything from subways to bicycle paths.

In previous highway legislation, a penny and a half was set aside, but only for mass transit. The new bill also lets states use federal highway land for building a route for a maglev or another high-speed train. A proviso that would have let states use discretionary dollars for tracks and vehicles for high-speed rail projects—seed money that could have attracted private capital for a maglev project—was dropped because of a jurisdictional dispute between two committees in the House of Representatives.

Some maglev zealots wonder whether the administration may be trying to kill the prototype program altogether by requesting that no money be appropriated for this no-holds-barred approach to technology development.

Indeed, the breakneck pace advocated by Moynihan may be a mistake. Transportation planners note parallels with other potentially revolutionary technology whose acceptance has been retarded because of an overstressed and unstructured development process. A "People Mover," a driverless rubber-tired train car that runs on a concrete guideway, was built in the 1970s in Morgantown, W.Va. Like maglev, it was billed as a substitute for the automobile. Costs on the project climbed from a low estimate of nearly $14 million to $118 million.

Too hasty a maglev effort—one that tries to resolve sticky design issues while readying a system for fare-paying passengers—could be programmed for failure. "If you try to put in a system at $1 billion and it costs $5 billion and nobody uses it, then it will be dead for the next 50 years," says Peter Benjamin, director of planning for the Washington Metropolitan Area Transit Authority who was involved with the Morgantown project when he worked for the Urban Mass Transportation Administration.

From Macy's to Bloomingdale's

A carefully paced maglev program will still require new thinking about how to use an air train. Those who favor the technology say it is wrong to conceive of a route map as if it were just a series of lines connecting Penn Station in New York to Union Station in Washington. As experts suggest, a maglev may fail to draw enough travelers away from the automobile and airplane.

One response to naysayers is a simple sentence written into Moynihan's maglev program that calls for developers to design and build a switch to divert vehicles to stations off of a main high-speed line, like an exit from an interstate. This would allow the technology to adapt to the reality that most passengers for this ground-based aircraft today are suburban dwellers, not denizens of the central cities. "A train does not meet the demographics of this country," Kolm says. "People no longer go from downtown to downtown—they go from shopping mall to shopping mall at 15-mile intervals."

An electronic switch that lets a vehicle move from one track to another at nearly full speed would improve on the more primitive electromechanical systems demonstrated so far by the Germans and Japanese. A sophisticated switch would allow an American maglev to fulfill the vision of a maglev as a hybrid, one that incorporates the character, not just of a plane or train, but of the automobile or jitney as well. These single-vehicle maglev cars carrying 140 passengers could follow one another within the space of a minute. East Sheboygan and its outskirts could be served as well as Milwaukee. Where access to the central city is required, the interstates, which Moynihan had wanted to route around the metropolis, may now redeem themselves by giving access to a downtown business district.

Argonne National Laboratory, with expertise in both magnetic technology and transportation planning, looked at maglev as a surrogate for the airplane. An Argonne study found that the technology could serve as a replacement for commercial airplanes on routes of less than 600 miles in length. "It's like flying a train next to the interstate," says Laurence E. Blow, a member of Argonne's technical staff. "You can't think about maglev from traditional planning assumptions."

Maglev is the kind of technology that holds an allure for elder statesmen, transportation visionaries and the type of scientist or engineer who, one might argue, should be eased into retirement with a tall stack of science fiction books. But as Moynihan points out, technological innovation is sorely needed to reverse flagging productivity in the transportation sector. Most of Ice Tea is a hodgepodge of programs that preserve or improve on the status quo by maintaining existing highways and helping to upgrade or build new transit systems. The risk-laden maglev program consumes about a half of a percentage point of the $151 billion allocated for the coming six years.

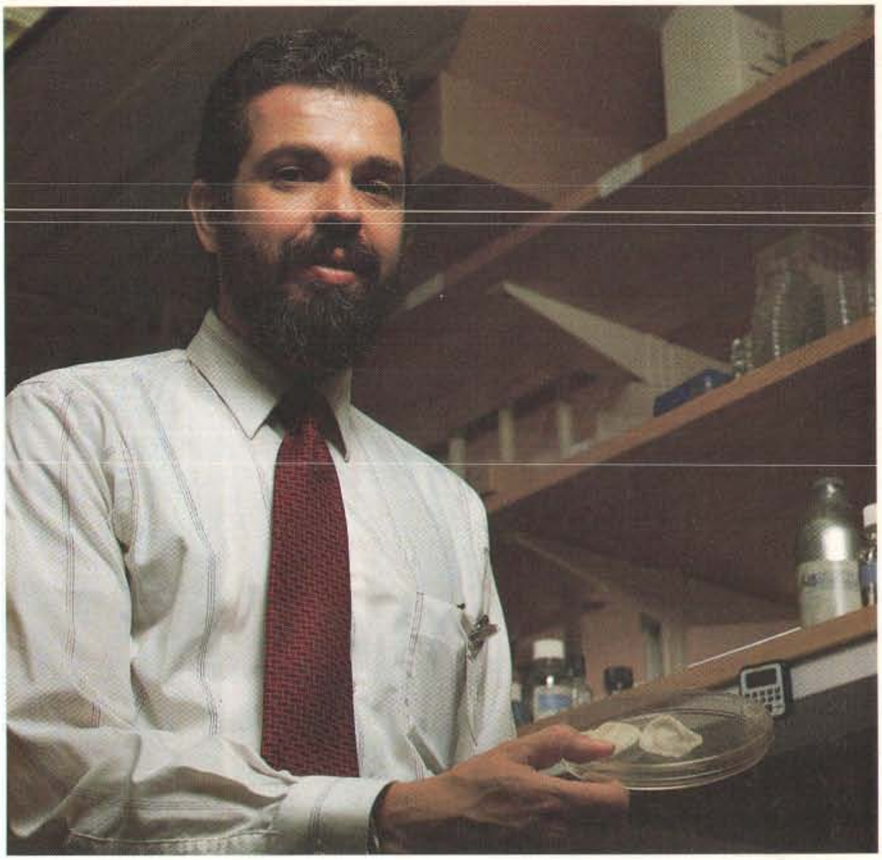
In recent years, Kolm has devoted his energies to a variety of projects, including the design of electromagnetic coil guns for the U.S. Army. Harnessing the inventive spirit of a Kolm for something besides an armor-piercing weapon against tanks might be worth a gamble that still amounts to about a third of the cost for a Seawolf submarine. Moreover, it might also prove that the federal government is capable of doing a little more than just trickling money for fixing bridges and potholes. "The country is being run by MBAs," Kolm grouses.

But maglev pioneer James Powell says it with a touch more reflection. "The whole spirit of America of the last century was one of people who were in love with doing new technology and people who did not let the difficulty and bigness of the step stand in the way of progress. People were fascinated by doing the Panama Canal. That sort of spirit has gone away. Something's wrong. For whatever reason, we've tended to lose the drive and willingness to do difficult things."

Research for this article was supplied by Deidre Berger in Frankfurt and Tom Koppel in Tokyo.

Material Help

Bioengineers produce versions of body tissues



CHARLES A. VACANTI displays ears molded from a biodegradable polymer designed to support the growth of cartilage cells. Photo: Stanley Rowin.

Since the late 1970s cellular biology, protein chemistry and bioengineering have opened insights into the ways cells congregate and function. These observations have provided the framework for a promising multidisciplinary technology known as tissue engineering. "Tissue engineering is like an orchestra," says Ron Cohen, vice president for medical affairs at Advanced Tissue Sciences (ATS) in La Jolla, Calif. "Everyone has been off studying their own instrument for 12 years, and now they're coming together to play Mozart."

Research on tissues such as skin, cartilage, bone, liver, intestine and ligament is proliferating. Harvard University, the Massachusetts Institute of Technology, the University of California at San Diego and the University of Michigan are among the most devoted academic proponents. ATS, which is now in the process of acquiring Neomorphics in Cambridge, Mass., is leading the way to market, followed by Organogenesis and BioSurface Technology, both in Cambridge, LifeCell in The Woodlands, Tex., and Skeletech in Cleveland, Ohio. More companies are expected to spring up.

"There is never enough tissue to go around in any kind of reconstructive surgery," points out Joseph P. Vacanti, director of the liver transplantation program at Children's Hospital in Boston. Cadavers cannot supply living material; animal tissue is rejected. So surgeons harvest what they can from other areas of the patient's body, filling in where appropriate with pieces of metal or plastic. A supply of cultured tissue could provide a way to heal children born with severe facial deformities or missing sections of intestinal tract; it could enable surgeons to treat people disfigured by burns, cancers and traumatic accidents and those whose joints have been eroded by hard use, arthritis or disease.

Tissue engineering might even help relieve the chronic shortage of transplantable organs. It was the scarcity of transplant livers, resulting in the death of young patients in his care, that caused Vacanti to think: "If we can't get the organs, let's make some." Today a liver tissue model he began developing in 1986 with Robert Langer, a chemical engineer at M.I.T. who specializes in biomaterials for drug delivery, has advanced enough to prolong the life of a rat for three months. Artificial livers suitable for testing in humans are still years off, Vacanti acknowledges, but Neomorphics, the tiny company that funds the two researchers, is creating less complex tissues such as cartilage that should be ready sooner.

Langer and Vacanti are not the only ones in the field whose initial goals have been reoriented toward more immediate objectives. "We decided to focus on growing the tissues with the broadest applications," says Cohen of ATS. At first the company, called Marrow-Tech on its founding in 1986, intended to grow marrow taken from people when they were young and healthy and store it for possible future genetic therapy. But ATS officials soon realized they would need to get a product to market sooner.

ATS deliberately started slowly. The firm's first product, a skin substitute trade-named Skin^2, was launched commercially in November 1990. It was designed to measure drug and cosmetic toxicity but was not expected to protect a body. "What we learned from our toxicology kits we rolled into a skin product suitable for burn victims," Cohen explains.

The resulting Dermagraft is a fine biodegradable mesh seeded with fibroblasts from the subsurface layer of skin known as the dermis. The artificial skin, which is being evaluated in 150 patients at 12 burn centers across the country, provides a healthy, living bed for epidermis taken from undamaged areas of the patient's body. The product is meant to obviate the deep graft patches now cut from normal skin. Cohen anticipates that Dermagraft may heal better than adult skin, because the polymer matrix is strewn with cells gathered from neonatal foreskins. Fetal tissue does not scar, he notes.

ATS is already testing Dermagraft in ulcers that arise in the legs because of poor circulation. The firm will soon as-

sess the product in other types of skin ulcers where abnormal subsurface tissue discourages normal epidermal cells from migrating into the wound. "Now we're taking what we learned with skin and applying it to cartilage, another tissue that the body has no capacity to regenerate," says Gail K. Naughton, the firm's chief operating officer.

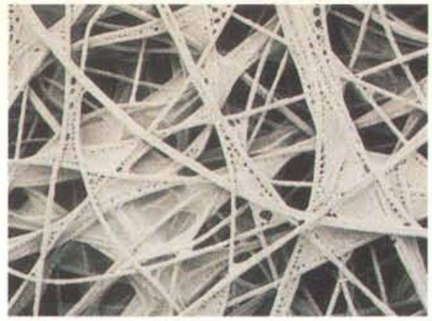
"The optimum replacement for cartilage—whether it's in chins or cheeks or joint linings—would be your own cartilage," says Charles A. Vacanti, Joseph's brother, who is by formal training an anesthesiologist and lately a researcher into the mechanisms of cartilage and bone growth. On the black counter of his laboratory at Children's Hospital sit molded polymer ears and disks ready for seeding with various cell types. Vacanti is collaborating on optimizing biomaterials for cell growth with Linda G. Cima, an associate professor at M.I.T., formerly of Langer's laboratory.

An orthopedic surgeon could scrape some cartilage from a knee, for instance, while doing the customary investigation with an arthroscope. The biopsied tissue would then be exposed to an enzyme to degrade the matrix that surrounds the cartilage cells, called chondrocytes. Once the cells are freed, they begin to multiply. Letting the cells adhere to a porous polymer that degrades as they grow encourages them to form a structural matrix of their own.

"If you put the proper amount of cells on the right kind of support, you end up with new cartilage that's just like the real thing," Vacanti says. The laboratory-grown substance he proffers is virtually identical to cartilage being removed from a fresh cow shoulder down the hall, he asserts—in chemical makeup as well as appearance. But verifying the material's performance in the body, such as its ability to bear weight, will be up to companies and the Food and Drug Administration, he cautions.

The most effective support yet used for growing cartilage is a tufted mass of polymer fibers that looks like a teased cotton ball. In rabbit knees where articular cartilage had been removed, the polymer enabled 80 percent regrowth of that tissue—a feat never before achieved without some means of containing the cells to prevent dispersal, Vacanti boasts. He thinks the polymer might have performed even better if he had given the cells more time to grow in the matrix before implanting it.

The rate at which cell-supporting polymers dissolve is one of the most important factors in the creation of substitute tissues. Ideally the matrix should degrade as cells grow into organized masses. To achieve this, Langer and his colleagues Antonio Micos and Denise Barrera are varying the quantities of ingredients in biomaterials such as polylactic glycolic acid (PLGA), which is used for sutures and drug delivery.



CARTILAGE CELLS fill in gaps between polymer fibers. Photo: C. A. Vacanti.

The idea is to get as many cells as possible into a synthetic support in order to obtain a huge surface to volume ratio, Langer explains. In this way, the substitute starts out more like normal tissue and results in speedier healing. "We used to use sheets and films of polymer, but you can't get many cells in," he reflects. "Then we thought about how nature solves the problem, and you can see the answer in seaweed, in trees, in lungs—it uses a three-dimensional branching structure to grow. We want to design polymers that way."

"Frankly," Langer observes, "the more the cell biologists learn, the better it is for us chemical engineers. Their work tells us what to build." For instance, adhesion molecules on cell surfaces have recently been discovered to help neighboring cells maintain their differentiated states. If mimics of these sequences could be attached to a polymer, they might keep cells functioning as intended. Barrera may have just made a significant step toward this goal by inserting into PLGA an amino acid that can serve as an attachment site for other molecules. If the new polymer is reproducible, she says, "we could graft on whatever biologists determine is important."

Tissue engineers are enthusiastic about the opportunities awaiting their technologies. Surgeons, the engineers say, could use computer-aided manufacturing programs to produce synthetic cellular supports in the operating room or select pre-molded shapes, such as ears or sockets or triangles that stack into a nose. Polymers could release growth factors as they degrade in order to boost dividing cells at critical times. Artificial tissues may even make what is already good that much better, by helping to support genetically modified cells in the style to which they are accustomed. *—Deborah Erickson*

Hopscotch Policy

Is the U.S. inching toward a strategy for technology?

Technology policy in the U.S. often resembles a game of hopscotch. Players toss down ideas—such as a new institute to channel technology funding—then try to hop around the ideological, bureaucratic and financial objections to advance their cause. In the past, these contests have often degenerated into squabbles over the rules of the game.

More recently, industry and various branches of the government seem increasingly eager to agree on a set of rules and play together. "The old animosity is beginning to break," declares Robert M. White, under secretary for technology at the Department of Commerce. In its place, he adds, is "a dramatic change in willingness of industry and government to work together and cooperate."

Indeed, those who have spent frustrating years trying to nudge a government that has eschewed "picking winners and losers" toward more support for technology development are beginning to sound distinctly optimistic. "I've been saying for about eight months that there's been a turnaround," declares Lewis M. Branscomb, who directs science, technology and public policy at the John F. Kennedy School of Government at Harvard University. "Right now government policy is changing faster empirically than the ideological debate is being resolved," he says.

Government leaders, mindful of ideological snares, are more circumspect. Eugene Wong, associate director for industrial technology at the Office of Science and Technology Policy (OSTP), characterizes the changes as "shifting the focus to areas where there is substantial agreement"—namely, to generic technologies. White concurs: "I don't think there has been an ideological change." Nevertheless, the speed at which technology becomes obsolete and the pace of competition have "forced us within the ideological framework to change how we behave," he says.

A flush of proposals and initiatives to extend such a framework are circulating around Capitol Hill. Among the most controversial is an idea advanced by a panel convened by the National Academies of Sciences and Engineering and chaired by former Secretary of Defense Harold Brown. After reciting the usual litany of government responsibilities (such as ensuring the economy is in working order), Brown's committee

lights a firecracker: the government should create a "quasipublic" entity that will invest in fledgling technology collaborations between the government and the private sector.

As described in Brown's report, this entity would receive a one-time injection of government financing—on the order of $5 billion—and then be left alone to choose which endeavors it should back. In this way, funding decisions will be insulated from the pork-barrel politicking that has often dictated which way federal monies flow, proponents argue. "This isn't a way to provide low-cost capital to firms making investments," insists John M. Deutch, a professor at the Massachusetts Institute of Technology and a member of the Brown commission.

Instead, Deutch argues, the government might consider nurturing generic technologies considered key to many other businesses but perhaps unprofitable themselves, such as flat-panel display screens. Yet even policy enthusiasts have raised questions about the breadth of the gulf between the taxpayers and the spenders. "If people spend $5 billion of the government's money, they've got to be accountable for it," Branscomb contends.

Another proposal, one that seems to reappear on the congressional agenda with the regularity of summer heat waves, is gaining more favor. It calls for reorganizing the Defense Advanced Research Projects Agency (DARPA) into a national research agency, dysphonically named NARPA. While continuing to explore technologies for the defense sector, NARPA would focus more intently on so-called dual-use technology and reach out to commercial firms that have seldom been defense contractors.

In past years, policymakers have questioned how readily DARPA could switch from serving one client—namely, the U.S. military—to satisfying the often contradictory needs of the civilian marketplace. "There's no support for a civilian DARPA in the administration," says Deborah L. Wince-Smith, assistant secretary for technology policy at the Department of Commerce.

Congress sees the issue differently. As the Defense Department itself retrenches for a period of peace, DARPA will have to direct its efforts increasingly toward civilian technologies, some experts suggest. "We are entering an era in which it is imperative that we link our competitive policy and our military policy," argues Senator Jeff Bingaman of New Mexico, one of the sponsors of the current NARPA bill.

Bingaman has also sponsored a program that has already survived several rounds of political wrangling. Called the Critical Technologies Institute, it aims to pick up where the OSTP's March 1991 report left off. That report offered a list highlighting 22 areas of technology considered vital for the health of the U.S. economy.

The Critical Technologies Institute will go a step further by providing a formal mechanism through which the government can regularly evaluate emerging technologies. The government will appoint a contractor to be responsible for charting technological road maps as well as for helping the OSTP coordinate research efforts endorsed by multiple government departments. One clear sign of changing times is that the go-ahead for the institute was clinched when Bingaman and the OSTP secured the backing of Richard Darman, the director of the Office of Management and Bud-

Biotechnology for Sale, Cheap

Has the U.S. biotechnology industry been selling off its seed corn? It certainly seems to be doing so in transactions with Japan, says a National Research Council study group. According to the council's recent report on U.S.-Japan linkages in biotechnology, 90 percent of technology transfer flows from the U.S. to Japan. Japan returns little but cash in the form of payments to struggling U.S. start-up companies.

Among the many reasons for the negative flow is Americans' confidence in their ability to stay on the cutting edge of technology, observes G. Steven Burrill, co-chairman of the NRC group. "Americans are always on to the next great whatever," he says. "We love the early curve of technology development." This tendency is undercutting U.S. economic competitiveness, Burrill cautions. While this country excels at basic science funded with tax dollars and invites others to look on, the Japanese focus on commercializing research where economic opportunities are already firmly established.

Beyond machismo, the report points to important structural differences in the two countries' corporate and financial systems. Virtually no small U.S.-style biotechnology companies exist in Japan. Instead pharmaceutical companies and other large manufacturing firms there are developing biotechnology to diversify into new areas and add value to existing products. These well-established corporations have been willing to make long-term investments in areas such as biosensors and bioprocessing.

U.S. biotechnology companies have been so eager to obtain capital that they have not asked for much more than cash, the NRC study states. Lopsided deals result, especially when the lure includes access to the coveted Japanese market. Yet "technology linkages," such as marketing, manufacturing and distribution agreements, are not the only way that technology travels overseas. Grants from a Japanese firm to an individual researcher at a U.S. university buy more than contract research. The Japanese are buying a calculus of research discoveries consolidated in any given researcher; the equations frequently include laboratory visits and training for foreign scientists. Even invitations to specialized conferences can serve as siphons, as do overnight faxes that tell all, the report warns.

"Instead of trying to create a level playing field, the more important question is how to compete and win," the NRC group concludes. One suggestion is to charge more. "If a superpower like Japan isn't willing to support more basic research at home, then why shouldn't they have to pay a premium to access it abroad?" challenges Susan Clymer, managing director of NichiBei Bio, a San Francisco–based consulting firm specializing in U.S.-Japanese biotechnology transactions. Clymer suggests that if American companies want Japan to bring something to the party, they should negotiate for access to the vast repositories of microbial cell lines Japan has collected from all over the world. These stores of fungi, plants and bacteria are what give Japan its strength in fermentation, she notes.

America's present lead in biotechnology is by no means a given. If one-way technology transfer continues and the U.S. fails to develop its ability to commercialize research globally, Japan will once again be poised to seize an enormously powerful and profitable technology, the NRC study cautions. The shift could be rendered permanent by the end of the decade, unless the U.S. learns how to get more in return for its research. —*Deborah Erickson*

get. Once Congress releases the $5 million earmarked for the institute, the program is expected to roll into action.

Also moving steadily ahead has been the Advanced Technology Program (ATP), administered by the Department of Commerce. When the effort was initiated by Congress in late 1988, the Bush administration extended only grudging acknowledgment—in the guise of a $10-million budget in fiscal 1990. But gradually the administration has proved more generous: it proposed awarding $68 million to the ATP in fiscal 1993. (Congress still hopes to boost that budget beyond $100 million.)

The Advanced Technology Program is spending its resources aggressively. In late April it awarded 27 grants for innovative technology projects, which will be jointly supported by industry. Those efforts range from a tiny startup working on fractal-based signal-processing techniques to a collaborative effort involving Honeywell and 3M aimed at developing neural-network computer programs for materials processing. "The interesting point is that when we map those awards against the OSTP's critical technologies, we find that we're building a portfolio that covers all those areas," White observes.

The National Institute of Standards and Technology is also playing an important role in pulling together an interagency technology road show—the National Technology Initiative. Scheduled to appear in 11 cities between February and July of this year, this multiagency extravaganza has given local businesses a peek into federal research projects and offered advice on how to establish partnerships without tripping over antitrust regulations.

Such programs point to a fundamental question that many believe is emerging from the debates over technology policy: How does technology move from the laboratory to the marketplace? "The issue isn't how to invent technology but how to race up the learning curve and implement technology," asserts Anthony Carnevale, president of the Institute for Workplace Learning in Alexandria, Va. The U.S. will move up this curve more rapidly when government and industry help and educate each other, rather than quarreling about the rules of the game, he adds. And increasingly, as the U.S. plays through this round of policy hopscotch, it has an eye on the international tournament going on in the adjacent lot. *—Elizabeth Corcoran*

Seven-League Boots

When an earlier generation yearned to "leap tall buildings in a single bound" like Superman, "P-F" Flyers offered springy sole inserts. In the age of Marathon Man, high-tech sneakers ballyhoo inflatable heels to recover energy that would otherwise merely heat the asphalt.

But if you really want to put bounce in your step, its inventors say, just strap yourself into a 55-pound contraption called the SpringWalker. It lengthens your stride, doubles your leverage and hoards twice as much energy as the next most efficient hopper. "We recover about 80 percent of the vertical energy," says John Dick, president of Applied Motion in Claremont, Calif. "Kangaroos recover about 40 percent, by far the best of any animal."

The goal is not necessarily to compete with pole vaulters without the aid of a pole. Instead, Dick says, the device will let a wearer lope through a marathon at speeds a sprinter might envy. "The present prototype is a little slow," Dick admits, adding that real performance will come when the company finishes its 25-pound sports model. "The long-range goal is a bipedal gait of 25 miles per hour"—Carl Lewis's average in the 100-meter dash.

Dick, a physicist who works on atomic clocks for the Jet Propulsion Laboratory, and his colleague Eric Edwards, an engineer at SRI International, designed the machine for fun. Now they hope for profits. They have patented the design in the U.S., have patents pending overseas and are talking with motorcycle makers, exercise-machine firms and the U.S. military. "If it is as much fun as it looks like, I take it there will be some money in it," Dick says.

The SpringWalker is a spidery exoskeleton of levers and stilts that straps to the torso and feet and suspends the wearer 18 inches above the ground. Pumping the feet stretches and relaxes a bungee cord mounted behind the operator's back. The shock-absorbing spring leaves joints unjarred, a feature that Dick says intrigues such exercise-machine companies as NordicTrack.

The pedals give a leverage of nearly two to one, lengthening the stride and slowing the bounce cycle. "You don't have complete freedom in the rate you're running," Dick says. "The device has a sort of natural resonance, like a trampoline." Because standing still without falling can be tricky, the company is designing mechanical toes to give the feet more purchase on the ground.

Then there is the possibility of installing a servomechanism. Applied Motion is working on a design that would add three parts of motor power to every part of leg power. Dick says the military and Yamaha have expressed interest in such a "man amplifier." With it, mere mortals might outrace Mercury himself. *—Philip E. Ross*

BOING!! goes the SpringWalker. Photo: Debra Lex.

Check It Out

A retina on a chip eyeballs bad paper

From The Gap to Acme Supermarket, store proprietors are the recipients of 13 billion or so personal checks every year, some of which are not worth the paper they are printed on. A simple technological fix for the problem would be a device that, like those that approve credit-card purchases, recognizes the account information on a check and automatically dials a data base to verify an account balance. But conventional magnetic check readers have difficulty recognizing account numbers because consumers mark and mutilate the documents with aplomb. So only a few retailers have adopted the technology.

A novel approach comes from two microchip pioneers, Carver Mead of the California Institute of Technology and Federico Faggin, who helped to design the first microprocessor when he was at Intel in the 1970s. Their invention, crafted along with Mead's former student Tim Allen, consists of a new type of image sensor and accompanying circuitry that mimics the functions of the human retina and the visual cortex. Mead and Faggin say this product is the first commercial application that embodies a so-called neural net on a chip. Most neural networks to date have been programmed in software because of the need to make frequent alterations to the processing elements and the lack of methods to design these specialized analog circuits.

The I-1000 chip will be manufactured by Synaptics, a company founded by Mead and Faggin in 1986. Synaptics developed the chip for a check reader made by VeriFone, a Redwood City, Calif., firm with $188 million in revenues that holds 64 percent of the market for the credit-card verification machines used by retailers. Officials of both companies hope the $250 average price of the check reader is sufficiently low—a fraction of the cost of many conventional check scanners—to foster acceptance by retailers. This type of machine might eventually be used to debit a retail customer's account automatically.

There are no established design techniques for combining thousands of analog transistors to process data all at the same time. Synaptics chose to reduce the risk of the project by mixing unconventional circuit elements with the building-block design methods espoused by Mead in his classic textbook on integrated-circuit design. This enabled the company to build a small, inexpensive chip that can be manufactured by most semiconductor fabrication facilities. "This is a path that could lead to massive deployment of neural networks," Faggin exclaims.

The device uses a "silicon retina," sensor circuitry that is fashioned to operate like the cone receptor cells of the retina. Like its natural counterpart, Mead and Faggin claim, their device can cope with differing light levels and variations in ink density as well as other kinds of environmental noise. Photodetectors in other optical scanners cannot compensate for light levels that may vary across the image, causing the misidentification of the bank and customer account codes on a check.

The image formed by the receptors of the silicon retina (a simpler version of the one described in the May 1991 issue of *Scientific American*) is transmitted to two neural networks on the same chip. Like the visual cortex, these networks can be "trained" to recognize the subtle differences between similar objects, a numeral 3 and an 8, for example. One of the networks locates a character; the other identifies it.

The recognition circuitry consists of 14 "neurons," one for each of the characters and numerals in the check code. Each neuron consists of 400 interconnected transistors that are the equivalent of the synapses in the brain. Within a neuron, each transistor processes one of the 400 picture elements from the image captured by the retina. The signal outputs for each transistor—the product of the brightness of each picture element and a specific value (a "synaptic weight") for the transistor—are summed. The neuron with the strongest cumulative signal "wins."

Like the human visual system, the neural network enables virtually instantaneous recognition. All 14 neurons process the captured image at the same time, an operation that takes no more than a microsecond. Also like the eye and brain, incomplete information can be processed. As long as the signal from a neuron reaches a preprogrammed threshold, the character is identified. A check number that has been inked over may produce a weaker signal, although one that is still sufficient for recognition. A conventional digital microprocessor then prepares the check data to be sent by a telecommunications link to the bank for verification.

The magnetic readers now offered to retailers have high error rates that result from misprinting or mishandling. For the most part, automatic check reading has been confined to the back offices of banks. There expensive scanners combine both optical (charge-coupled devices) and magnetic-sensing elements. The two sensors are often connected to signal-processing computers and then larger computers that actually handle the account transactions.

Synaptics chose a simple application

*MICROELECTRONICS PIONEERS Carver Mead and Federico Faggin are commercializing a neural-network chip. Photo: Fran Ortiz/*San Francisco Examiner.

instead of trying to tackle a more ambitious problem such as making a vision chip for a robot. In a similar way, Faggin designed the first microprocessor for a desktop calculator, not a supercomputer. "I believe in the principle of walking before you run," he says.

The company is studying the design of a chip that can recognize counterfeit and different denominations of currency. Another deceptively simple chip under development can help discern type-printed addresses on envelopes. The U.S. Postal Service may be an eager buyer: more than 40 percent of the letters that move through its automatic address readers cannot be processed by the machines. *—Gary Stix*

Doomsday Diagnostic?

A precursor protein may predict the risk for Alzheimer's disease

Until there is a cure, does anyone really want to know whether he or she will someday get Alzheimer's disease? Scientists at SIBIA, the corporate spin-off of the Salk Institute for Biotechnology in La Jolla, Calif., are betting the answer is yes. They are preparing to commercialize the first predictive diagnostic test for the debilitating, fatal disease. "If it's going to afflict the breadwinner in the family, people might want to know," declares Steven L. Wagner, a SIBIA biochemist.

At present, a firm diagnosis of Alzheimer's disease can be made only after the fact, by autopsy: the brain reveals characteristic plaques, or lesions, of amyloid beta-protein. But SIBIA's test may work before irreversible symptoms appear. It can also distinguish Alzheimer's disease from other kinds of dementia, possibly helping to redefine treatment. The test could also be used to monitor the response of Alzheimer's patients to experimental drugs as well as the progression of the disease. "This is not just a doomsday diagnostic," Wagner asserts.

The test, which requires withdrawing cerebrospinal fluid, uses a highly specific monoclonal antibody to assay levels of amyloid beta-protein precursor (APP), a substance that gives rise to the protein plaques found in the brains of Alzheimer's victims. SIBIA and its collaborators have recently published a number of papers in such journals as the *Proceedings of the National Academy of Sciences, Annals of Neurology* and the *Lancet*. The articles all conclude that Alzheimer's disease is associated with decreased levels of APP. "This isn't just any biological marker. It is the protein critically involved with the disease," says William E. Van Nostrand, who developed the antibody with Wagner at the University of California at Irvine.

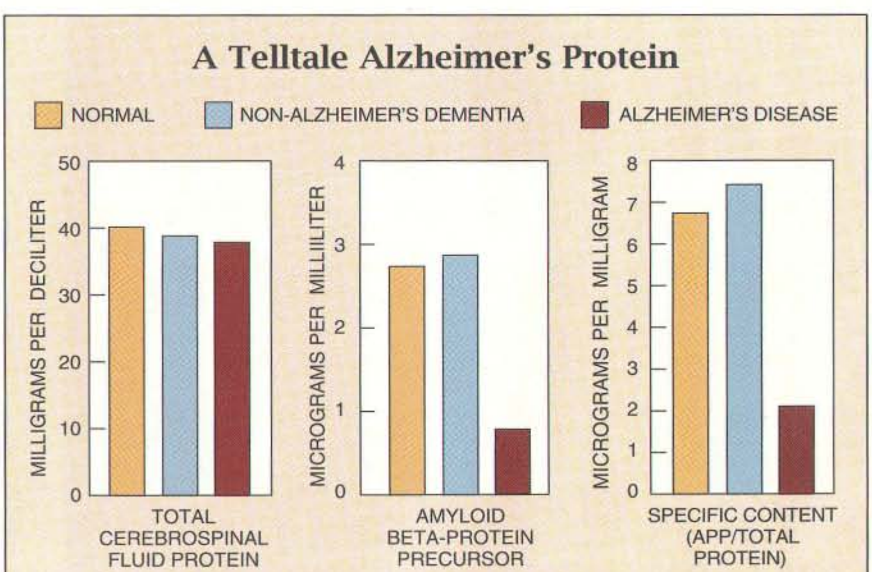
Scientists at U.C.I. and their colleagues at the Free University Hospital in Amsterdam, the Netherlands, used the antibody assay to test APP levels in 16 healthy patients, 13 probable Alzheimer's patients and 18 others with unrelated dementia. APP levels in the Alzheimer's patients were about 3.5 times lower. The researchers contend that their findings indicate abnormal processing of the protein that forms amyloid deposits. (The association does nothing to clarify the ongoing chicken-and-egg debate as to whether amyloid causes Alzheimer's or results from it.)

Other research also demonstrates a correlation between the disease and decreased levels of APP, the SIBIA scientists say. For instance, in one study the company conducted at the University of California at Los Angeles, "we were able to pick the severely affected guy out, presymptomatically," Wagner says. Four patients at the John Douglas French Center in Los Alamitos, Calif., had spinal fluid withdrawn before taking an experimental drug; their levels of APP were 2.4, 2.3, 1.4 and 0.1 micrograms per milliliter. "The one with the lowest reading was in the best shape at the time the trial began," Wagner recalls. "Now he's the most afflicted."

The company has even done some postmortem sleuthing to confirm the protein's role. "We went back and looked at spinal fluid taken from patients three to six years before they were autopsy confirmed as having Alzheimer's," Wagner explains. Few centers had bothered to store the fluid, he recalls, but Mount Sinai Hospital in the Bronx and the University of Washington produced some from the deep freeze. SIBIA analyzed the samples, and the centers matched their findings with pathology reports. The lowest levels of APP belonged to patients who had the highest levels of amyloid plaques throughout their brains, especially in the blood vessels there. "We think this particular lesion may have the most dramatic correlation with our test," says Van Nostrand, who now consults for SIBIA.

SIBIA anticipates bringing a commercial version of the test to market by the end of next year, initially via the 25 to 30 Alzheimer's disease centers across the nation. Preliminary versions are currently being evaluated at several of these. SIBIA's test is never likely to be a doctor's office item, but it could certainly be done in a hospital, Wagner says. "Cerebrospinal fluid is taken much more often than you'd think," he explains, to determine whether patients undergoing neurological exams have multiple sclerosis, an infection or some kind of autoimmune disorder.

The company will not yet comment on the price it will ask for the test. "The value is great, especially in terms of what it takes to get a result by other means," Wagner says. In the meantime, SIBIA plans to use the assay to evaluate its own nascent drug for Alzheimer's disease. It would be more convenient if scientists could develop a blood test for the disease. But that is unlikely because the disorder is confined to the central nervous system, Wagner points out. He shrugs: "We know the limits of our test, but for now there's nothing better." *—Deborah Erickson*



SOURCE: SIBIA

THE ANALYTICAL ECONOMIST

Experimenting with the Invisible Hand

One of the prime tenets of most economists is that markets can efficiently set prices that match the producer's costs with the benefits the consumers receive. But what is a market anyway? Specialists have identified any number of different methods for buying and selling.

These forms of ritualized haggling run a bewildering gamut. There are single oral auctions in which hopeful buyers bid up the price until only one is left and double oral auctions in which buyers and sellers negotiate for a price. Then there are English and Dutch auctions in which the price rises or falls with the tick of a clock, as well as sealed bids, posted prices or prices set by the last bidder to drop out.

It turns out that this is no mere taxonomic exercise. The efficiency of a market depends directly on the method it employs. Buyers in sealed-bid auctions, for example, tend to pay inflated prices because they have no idea what others will bid, and the highest estimate wins. On the other hand, sealed-bid contract competitions give jobs to the lowest bidders, even if their prices are unrealistically low.

Experimental market economists have begun to study the effects of different market institutions in a straightforward way that might seem alien to economists weaned on huge compilations of national data and intricate statistical analyses. These mavericks recruit people to trade in simulated markets, offering them the incentive of the profits on their ersatz transactions. Vernon L. Smith of the University of Arizona pioneered the interactive computer-based market, which allows programmers to look over traders' shoulders and to change market rules at whim.

There is no formal theory of how market institutions influence efficiency, says Raymond C. Battalio of Texas A&M University; inventing one will probably require cooperation among game theorists and psychologists as well as economists. The good news, however, is that theoreticians are beginning to take notice that it is possible to do controlled economic experiments. As in other sciences, theorists and experimentalists must cooperate to make sure that experiments are actually designed to test interesting questions, Battalio notes.

Some results have already begun to come in, and a few theoretical predictions have been fulfilled. For instance, Nobel Prize winner Ronald Coase has argued that companies hire employees on long-term contracts because it would be too expensive for both sides to negotiate prices on a piecework basis—$50 for a memo, $25,000 for a strategic plan and so on. One-shot contracts are economically efficient only when the value to both parties of the item being bought or sold is clearly defined. Battalio created a simulation in which traders could decide during each round whether they wanted to make one-shot or long-term deals.

When traders had full information about one another's costs and benefits, they bought and sold at arm's length, but when Battalio changed market conditions midstream to reduce their information, they chose employer-employee rules instead. If he started the traders out with minimal information and later gave them full knowledge, traders once again picked the most efficient market institution for each case.

The economic efficiency of a market depends on appropriate trading mechanisms.

Battalio also found that two market mechanisms may appear identical under conventional analysis but differ substantially in their practical results. He ran a two-part experiment in which players had to buy the right to be allowed to take part in a second set of transactions and found that the economic efficiency of the second set of transactions depended on the market institution used for the buy-in. If entry to the second round was free, traders made mutually unprofitable decisions in that round, as they did when Battalio used a Dutch auction (in which each tick of the auction clock lowers the price until someone makes a bid) to set the entry price. Then Battalio used an English auction, in which each tick sends the price higher, and bidders drop out until only one is left. Although theory predicts that the two mechanisms should be identical in their effects, those who made it to the second round under the English auction made more mutually profitable choices.

In addition to testing economic theories, such laboratory experiments can help shape real markets. David P. Porter of the California Institute of Technology has built an experimental market for scientists at the National Aeronautics and Space Administration working on the *Cassini* space probe to Saturn. Researchers designing different experiments can trade allocations of funding, mass, power consumption and data-transmission requirements among themselves to optimize the total performance of the spacecraft.

Porter tested a number of different market institutions before settling on a "smart barter" system in which the computer automatically looks for multiway transactions. In the system, for example, a researcher working on one instrument with a surplus mass allocation but not enough cash can strike simultaneous deals with a second offering mass in return for power and a third offering power in return for funding.

Another project of Porter's is a simulated market for air pollution rights, which will soon have a real counterpart when southern California implements tough clean air requirements. Contrary to predictions, Porter found that issuing "emission shares" to each polluter and allowing them to trade as needed—so that many shares would never trade at all—was just as efficient economically as a "revenue-neutral auction" in which polluters would have to bid for all the shares they needed and then split the auction proceeds.

As the cost of the personal computers required to run them continues to fall, new experimental market facilities are suddenly springing up. Smith cites laboratories in York, Bonn, Frankfurt, Barcelona, Madrid and Grenoble. He is also working with groups in Poland and Hungary to set up laboratories that will train people who have never traded before to participate in markets that do not yet exist. He speculates that the formerly socialist countries may adopt advanced computer-based trading schemes faster than the West because they have no entrenched institutions to slow them down.—*Paul Wallich*

MATHEMATICAL RECREATIONS

by Ian Stewart

The Interplanetary Olympics

Claudia Astrochild had traveled for three miserable months to the planet Mercury so she could preside over the annual meeting of the Olympic Committee. She detested space flight, and even now she felt queasy in the low-gravity environment of the planet. As president of the Olympic Committee, she felt an obligation to meet the planetary representatives face-to-face. But she could not remember why she had agreed to hold the meeting at the Mercury City Convention Center. As she walked into the meeting hall, a crowd of reporters attacked her.

"President Astrochild," one of the journalists shouted, "why is your organization spending half its budget to hold this meeting on Mercury?"

"As I have announced, we plan to hold the very first Interplanetary Olympics in 2092. It will be a historic occasion, but it also presents unprecedented logistic problems. I am meeting with all of the planetary representatives here on Mercury to sort through the issues."

"What sort of problems?" another reporter yelled.

"All the events will occur at the same time on all the planets and the moon because we simply cannot transport millions of athletes and spectators around the Solar System. The committee must find ways to coordinate the activities. I should mention that the Interspace Broadcasting Company has already paid our organization $1 trillion to televise the events on each planet."

"President Astrochild, do you have any ideas about how the event will be coordinated?"

"I do. On each world, athletes will perform under the supervision of local officials. The overall results will be collated centrally. We will award medals to the competitors who put in the best three performances. The same principles will apply for races, but competitors will run against the clock."

"Ms. Astrochild, how can there be any meaningful competition if conditions differ so much on each planet?"

"That issue is being discussed. I thank you for your questions, and I would like to thank the people of Mercury, wholeheartedly, for all the courtesies they have extended to the Olympic Committee."

Pushing her way through the crowd, Astrochild entered the meeting room. She asked the delegates to be seated and announced the first item on her agenda.

The Lunar delegate, Wilma Craterson, raised her hand. "Ms. President, I wish to raise an objection. I believe that Jovian athletes will have an unfair advantage in weight lifting because of the planet's great diameter."

"What?" yelled Thomas Headsquat, the representative from Jupiter. "That's ridiculous!"

"It is well established that weight lifters perform best when the event is held at high altitude," Craterson explained. "Since altitude can be defined by the distance from the planet's center and since the diameter of Jupiter is greater than any other planet, the Jovians will have an enormous advantage."

"Good heavens, really?" Astrochild stammered. "That's amazing."

Headsquat rose angrily to his feet. "Has it occurred to the Lunar delegate that Jupiter's greater gravity puts its team at a considerable disadvantage in weight lifting? Gravitational forces make much more of a difference than the effects of altitude. In any case, altitude should be measured from sea level, not the center of the planet."

"Jupiter hasn't even got a sea," Craterson pointed out.

"Neither has the Moon."

"It does so. The Sea of Tranquillity!"

"My dear colleagues," Astrochild pleaded. "Let me emphasize that last word—tranquillity. We must discuss these issues amicably to demonstrate to the entire Solar System our commitment to good sportsmanship."

"And sportswomanship," Craterson remarked.

"That's all very well," the delegate from Saturn said, "but you do know that strong gravitational forces make it a lot harder to jump long distances on Saturn than on the Moon."

Craterson jumped to her feet. "I move that the results of the weight-lifting competition should be adjusted to compensate for the effect of planetary diameter."

"Seconded," barked the delegate from Pluto.

"But that's absolutely idiotic," Headsquat screamed. "The Moon has a small diameter and low gravity. Anyone can lift tons on the Moon, and now you want to distort the figures even further in your favor!"

Astrochild sighed. "Fellow delegates, we all realize that conditions differ greatly across the Solar System. I believe we should investigate how those conditions influence athletic performance, and if need be, we should make adjustments to the results on each planet. We cannot possibly tackle all the issues here. I propose that we set up a subcommittee to discuss the issues. The group should clearly involve the delegates from the Moon and Jupiter, and I propose that the delegate from Venus should represent the medium-size planets. And assuming we can agree on that, I will now move on to the much more serious question. How large should our sponsors' logos appear on the Olympic uniforms?"

A few hours later Craterson, Headsquat and Alan Blandy, the representative from Venus, convened the first meeting of the Subcommittee on Interplanetary Athletic Fairness.

"By majority vote, we are agreed that diameter is not an important influence on weight lifting," Blandy said, receiving a nod from Headsquat and a hostile stare from Craterson. "Moreover, we are agreed that the main influence is gravity and that all masses should be multiplied by the local gravitational force."

"That's what weight is," Headsquat pointed out. "Mass times gravity. If gravity didn't matter, the sport would have been called mass lifting."

"That may be true," Craterson commented. "But the sport also requires that the athlete lift the weight up as rapidly as possible. And there it's a question of an impulse rather than a sustained force. An impulse is a change in momentum, and momentum depends on mass, not weight."

"We will note your dissenting view in the final report," Blandy said. "Now to the second item on our agenda: track events."

"That's very tricky if you get too detailed," Headsquat cautioned. "The best we can do is analyze a plausible model. I think there's a simplification that will suffice for all events except perhaps the short races. For now, we need only consider the horizontal component of

the athlete's velocity. When an athlete runs, chemical energy in the body is converted into kinetic energy of motion; the important component is that which propels the athlete forward. The chemical energy E that can be generated by the average athlete is the same on all member planets. Moreover, the kinetic energy is equal to one half the mass times the square of the velocity, that is, $E = \frac{1}{2}mv^2$ where m is the mass and v is the velocity. Therefore, the velocity is equal to $\sqrt{2E/m}$, which is independent of the force of gravity. In short, there is no need to adjust for gravity in track events."

"I guess," Craterson responded. "But running involves some up-and-down motion, too, and that movement depends on gravity. Anyway, you're neglecting the initial acceleration phase."

"Which is why I think we need to examine the shorter events more carefully. But not right now. As for vertical motion, I think it's up to the athletes to develop techniques to minimize it."

They provisionally agreed to recommend no compensation for the longer track events. Blandy then suggested that they discuss the high jump.

"That certainly depends on gravity," Craterson declared.

"Yes, I agree," Headsquat said. "I even have a theory about how gravity influences the high jump."

"Doesn't surprise me," Craterson snapped.

"Think of a jumper whose mass is m," Headsquat said. "The athlete can leap vertically at an initial velocity v, which is related to the impulsive force exerted by the muscles and is independent of gravity. Once again, the initial kinetic energy E equals $\frac{1}{2}mv^2$. If the jumper rises to height h, then kinetic energy is converted to potential energy, which equals the mass times the force of gravity times the height, or mgh. So $mgh = \frac{1}{2}mv^2$, and therefore $h = \frac{v^2}{2g}$. That depends on g all right. The bigger g, the lower the jump."

"Can I suggest one small correction?" Blandy asked. "What you have calculated is the height to which the athlete's center of mass can rise. Yet high jumpers can bend their bodies in such a way that they clear the bar but their center of mass remains somewhat below the bar. To allow for that, I suggest we apply a correction of 0.2 meter."

"You've forgotten something else," Craterson said. "The athlete's center of mass doesn't start off at ground level. It's about a meter high."

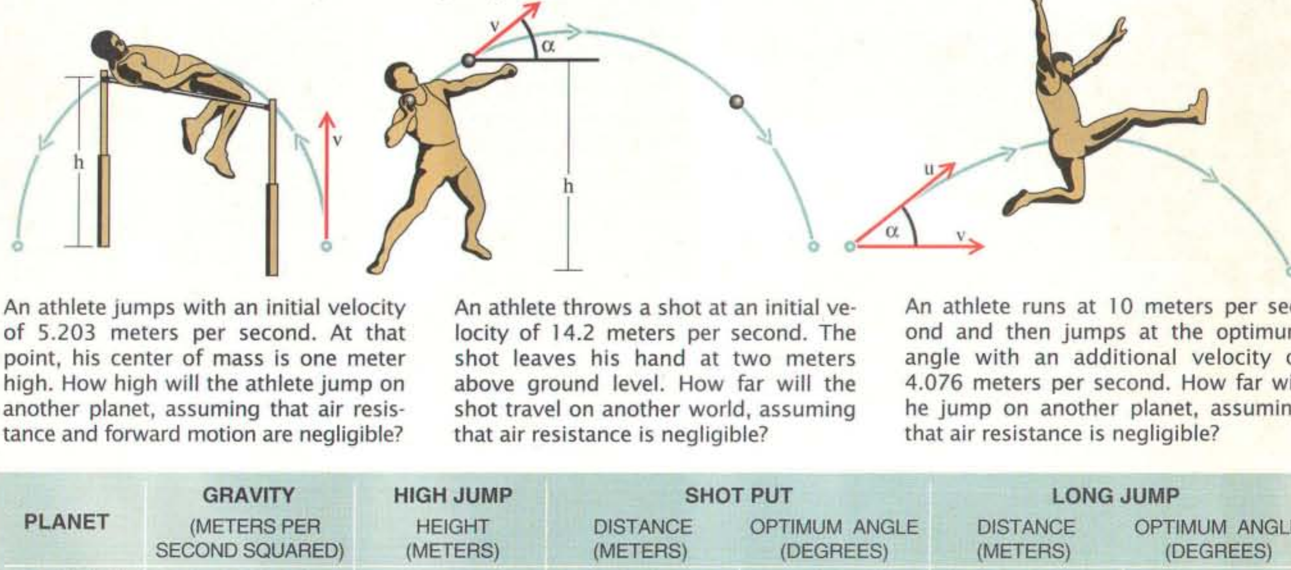
"I propose," Headsquat said, "that we apply a correction of precisely one meter. Then the actual height jumped will be $h = 1 + \frac{v^2}{2g}$."

"Can we draw up some comparisons based on that formula?" Headsquat asked.

"Sure," Blandy replied. "Let's see. About a century ago, in 1988, the Olympic record for the men's high jump on Earth was 2.38 meters, which was set by Gennadiy Avdeyenko of the Soviet Union. Louise Ritter of the U.S. jumped 2.03 meters to set the women's high-jump record. Using the formula, I calculate that Avdeyenko's vertical velocity was 5.203 meters per second, and Ritter's was 4.495 meters per second. Had they made the same jump on the Moon, they would have reached 9.36 and 7.24 meters, respectively. And on Jupiter, only 1.51 meters and 1.38 meters."

"No fair," Craterson cried. "You mean the athletes on the Moon must leap over a two-story building, whereas the

Olympic Feats on Other Worlds



An athlete jumps with an initial velocity of 5.203 meters per second. At that point, his center of mass is one meter high. How high will the athlete jump on another planet, assuming that air resistance and forward motion are negligible?

An athlete throws a shot at an initial velocity of 14.2 meters per second. The shot leaves his hand at two meters above ground level. How far will the shot travel on another world, assuming that air resistance is negligible?

An athlete runs at 10 meters per second and then jumps at the optimum angle with an additional velocity of 4.076 meters per second. How far will he jump on another planet, assuming that air resistance is negligible?

PLANET	GRAVITY (METERS PER SECOND SQUARED)	HIGH JUMP HEIGHT (METERS)	SHOT PUT DISTANCE (METERS)	SHOT PUT OPTIMUM ANGLE (DEGREES)	LONG JUMP DISTANCE (METERS)	LONG JUMP OPTIMUM ANGLE (DEGREES)
MERCURY	3.70	4.66	56.46	43.99	23.60	71.17
VENUS	8.85	2.53	24.70	42.69	9.87	71.17
EARTH	9.81	2.38	22.47	42.46	8.90	71.17
MOON	1.62	9.36	126.45	44.50	53.89	71.17
MARS	3.72	4.61	56.17	43.89	23.47	71.17
JUPITER	26.39	1.51	9.43	39.01	3.31	71.17
SATURN	11.67	2.16	19.17	42.02	7.48	71.17
URANUS	11.48	2.18	19.46	42.07	7.61	71.17
NEPTUNE	11.97	2.13	18.74	41.95	7.29	71.17
PLUTO	1.96	7.91	104.86	44.45	44.54	71.17

Jovian athletes need only jump onto their desks."

"Can't argue with the numbers," Blandy chuckled.

"Let's move onto the shot put," Headsquat demanded.

"Right," Blandy said. "Now, it's well known that in order to maximize the distance traveled, the shot should be launched at an angle of 45 degrees. That should help keep the analysis simple."

"It would—except it's wrong," Craterson sneered. "The 45-degree rule is valid only for projectiles launched from ground level. And shots are launched from approximately shoulder height [*see illustration on opposite page*]. Assume the shot is released from height h above ground at which point its velocity is v at an angle α to the horizontal. The distance traveled horizontally is then equal to the time in the air multiplied by the horizontal component of the velocity, that is, t times $v\cos(\alpha)$. (The vertical component of the velocity is $v\sin(\alpha)$.) Furthermore, at any point, the height of the shot equals the initial height plus the contribution from the throw minus the contribution from gravity. The height of the shot is therefore $h + vt\sin(\alpha) - {}^1/_2\,gt^2$. If we ignore air resistance—"

Craterson paused for a moment to fiddle with her wrist computer. "I calculate that the maximum distance traveled is $\frac{v}{g}\sqrt{v^2+2gh}$, and to achieve that the angle has to satisfy the equation

$$\sin(\alpha) = \frac{v}{\sqrt{2(v^2+gh)}}."$$

"Let me check my almanac," Blandy said. "Ulf Timmermann of the former German Democratic Republic set the men's Olympic record by throwing a 7.25-kilogram shot a distance of 22.47 meters. The women's record holder, Ilona Slupianek of the same country, lofted a four-kilogram shot for 22.41 meters. If a shot was thrown for 22.47 meters and was released at the optimum angle at about two meters off the ground, then its initial velocity was 14.2 meters per second. The optimum angle at that speed was therefore 42.46 degrees."

"And on Jupiter?" Headsquat asked.

"Assuming the same initial velocity, the optimum angle would have been 39.01 degrees, and the distance would have been a mere 9.43 meters. On the Moon, in contrast, the angle would have been 44.50 degrees and the distance an impressive 126.45 meters."

"What about corrections for the effect of altitude?" Craterson said stubbornly. The others glared at her.

"I propose we next consider the long jump," Headsquat said, "though we call it the horizontal jump, because on Jupiter it isn't actually very long."

"Don't forget this time that the athlete's center of mass is not at ground level," Craterson interjected.

"I don't think we need to worry about that. When the athlete's feet hit the sand, the center of mass is almost at the same height as it was when his feet left the ground."

"But when jumpers land," Craterson insisted, "they stick their feet way out in front and slide into the sand."

"By Jupiter, it's only an approximation, Craterson!"

"Headsquat's right," Blandy remarked. "We should start with a simple model. Let's assume the height of the athlete's center of mass is the same at the beginning and end of the jump. We can always think about complications later. Say the athlete is moving forward at velocity v at the beginning of the jump. The jumper also exerts an impulse that imparts an additional velocity u at an angle α to the horizontal. I then calculate that the optimum angle α—oh, heck, which button do you push on this thing? The result must satisfy the equation

$$\cos(\alpha) = \frac{-v+\sqrt{v^2+8u^2}}{4u},$$

which, curiously, is independent of gravity. The maximum distance traveled is

$$\frac{2u\sin(\alpha)\,[v+u\cos(\alpha)]}{g},$$

and that does depend on gravity."

Craterson peeked into the almanac. "In 1988 Bob Beamon of the U.S. held the Olympic record in the men's long jump, 8.90 meters, and Jackie Joyner-Kersee of the U.S. set the women's record, 7.40 meters. Assuming Beamon approached the takeoff point at 10 meters per second and jumped at the optimum angle, his initial velocity u works out as 4.076 meters per second. The optimum angle is 71.17 degrees."

"That can't be right," Headsquat said. "Long jumpers don't look like they're moving at an angle of 71 degrees."

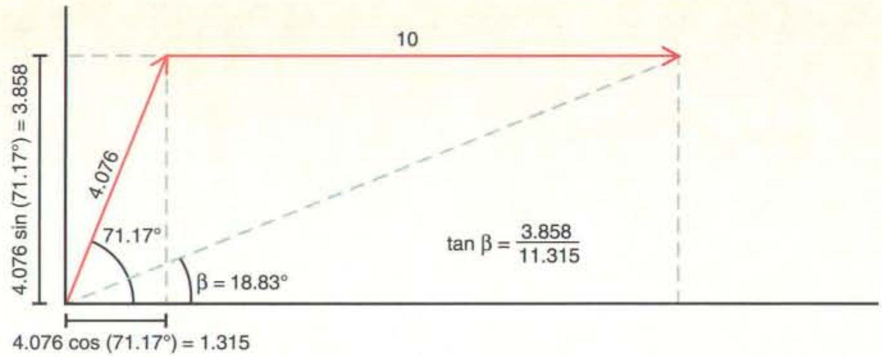
"That's because they're not."

"But you just said—"

"It's 71 degrees relative to the athlete. But the athlete is moving forward at 10 meters per second. Beamon's actual launch angle, as viewed by a spectator, would have been 18.83 degrees [*see illustration above*]. Long jumpers follow quite flat trajectories because most of their effort goes into forward motion."

After months of discussion, Craterson, Blandy and Headsquat found a way to adjust the results of all the Olympic events, and they were confident that their calculations would compensate completely for the effects of gravity. Indeed, the First Interplanetary Olympics was a great success, with two exceptions. The Lunar team won every event, and Craterson mysteriously disappeared. The incidents baffled Blandy and Headsquat until they checked the atmospheric control center in the Lunar Olympic Stadium.

"Welcome to the control center," a computer voice cheered. "I have been programmed to simulate Earth's atmosphere at an altitude of 1,000 meters above sea level. May I help you?"



***APPARENT ANGLE** of a long jump differs from the angle at which an athlete jumps. Consider an athlete who runs at 10 meters per second and then jumps with an additional velocity of 4.076 meters per second at an angle of 71.17 degrees. The jump adds 4.076 times cos(71.17) meters per second to the jumper's velocity in the horizontal direction. The jump also gives the athlete a velocity in the vertical direction, a component equal to 4.076 times sin(71.17) meters per second. The apparent angle is therefore 18.83 degrees, as calculated above.*

FURTHER READING

THE THEORY OF CLASSICAL DYNAMICS. J. B. Griffiths. Cambridge University Press, 1985.

THE MECHANICS OF ATHLETICS. Geoffrey Dyson. Holmes and Meier, 1986.

A FIRST COURSE IN MECHANICS. Mary Lunn. Oxford University Press, 1991.

BOOK REVIEWS by Philip Morrison

Atoms on View

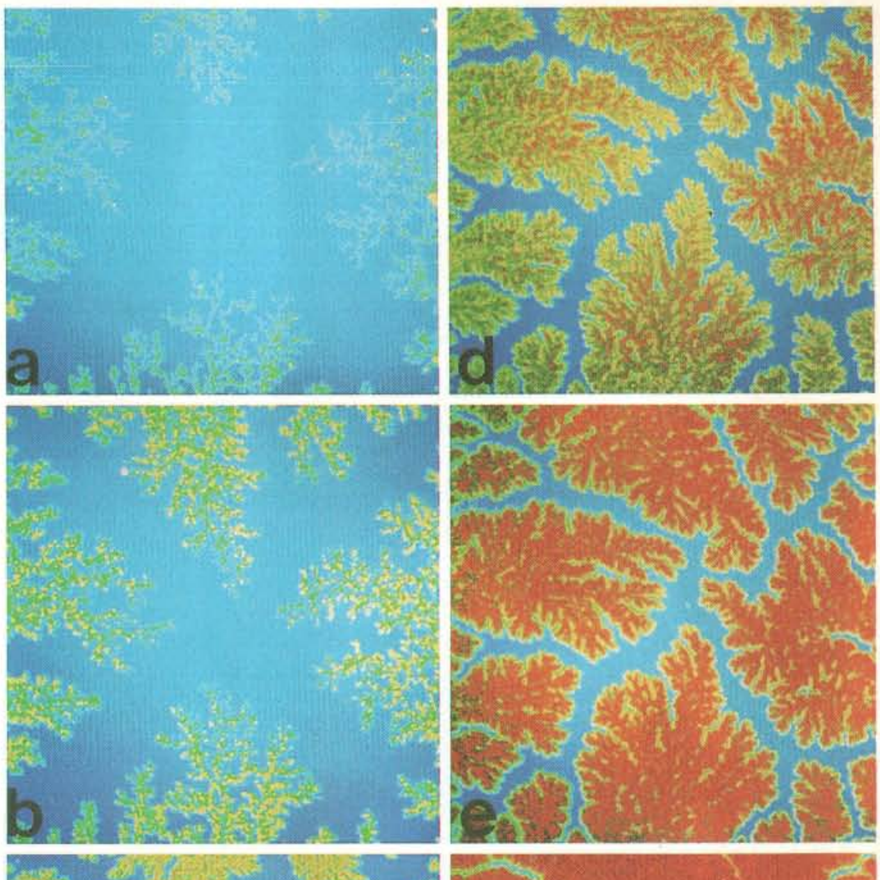
IMAGES OF MATERIALS, edited by David B. Williams, Alan R. Pelton and Ronald Gronsky. Oxford University Press, 1991 ($75).

Today's fine-structure imaging of the metals, polymers, ceramics and semiconductors of our technology can hardly be called microscopy. Call it rather the microscopies. The demanding art is remarkably many-sided. Each of the dozen chapters in this book is a state-of-the-art overview of one pictorial discipline by an expert author or a small group. Their welcome practice is to break off some highly technical explication full of acronyms and urge the nonspecialist reader to attend instead to the images overleaf. The level of the text remains demanding; this is a resource for materials scientists, but opened for others through its uncommon visuals.

Virtuosity abounds. Atomic resolution is neither commonplace nor yet extraordinary. A grand transmission electron microscope (TEM) at Berkeley is especially stable, both mechanically and electronically. Lens aberrations reduced by an order of magnitude, it is routinely able to resolve at the atomic level. A whole series of TEM photographs shows the local atomic defects that make ideal lattices interesting—dislocations, grain boundaries and the like. Perhaps the best image is one of a graphite film seen in phase contrast. A couple of dozen roughly square crystallites fill the scene, each a stack of 10 or 20 well-resolved, neatly parallel planes of carbon atoms, neighboring stacks randomly aligned.

Other pages document an active trip into the atomic world through a much simpler instrument. A photograph shows a tiny portion of a nickel-molybdenum alloy. The field is very small; the image displays only 100 or 200 atoms, a few atomic layers on a tiny sharp needle point of tough metal. An oddly bright spot appears in the image among an orderly set of 15 or 20 identical atoms. It must be a misplaced atom in the ordered layer, but how can we be sure? A controlled nanosecond pulse at high voltage removes a few atoms, and the bright interloper goes off among them. But the experimenters have first positioned the movable small opening of a time-of-flight mass spectrometer at just the right spot behind the glowing screen. The bright atom flew off into the spectrometer. Its arrival was duly detected down the line by the single-atom detector there. The detector is a sensitive double-channel plate, effectively a bundle of tiny tubes, within which the entering atom collides with the tube walls, to release a fast-multiplying avalanche of secondary electrons. The flight time signaled was conclusive: that bright interloper was a single molybdenum atom ushered out of its wrong seat.

This is the modern form of the wonderful field-ion microscope first shown



MICROGRAPHS monitor the growth of gallium oxide **(green, yellow and red, in order of increasing intensity)** *on the clean surface of liquid gallium* **(which appears blue).** *The oxide extends in a fractal pattern from the edges, thickening* **(appearing redder),** *until it eventually covers the entire surface.*

us in the early 1950s by Erwin Müller. A sharp metal tip has a point radius of a few hundred atoms, pointed toward a flat channel plate at a distance of two inches, a million times the tip radius. The applied voltages ionize some of the gas atoms in the fine but imperfect vacuum just outside the tip. The ions are drawn straight outward to the phosphor screen, where they paint out at millionfold magnification the protruding spots—the few dozen atoms—on the tip. (Nowadays another channel plate catches each ion first, to intensify the screen glow and allow sampling for precise identification of each spot.) Virtuosi!

Those redoubtable atom resolvers are image-forming devices; they map a steadily illuminated field all at once, like any light microscope. Today in most microscopes a tight scanning beam paints the image spot by spot as on a video screen. Spatial resolution depends on the beam size; magnification on the video display size; image contrast on the physical nature of the beam and on what properties are measured and displayed. Scanning beams do not aim at atomic resolution but rather provide a whole atlas of analytic mappings by exploiting the various physical returns a choice among beams can induce. Electrons, ions, ultrasound are all now in use.

In this decade the scanning tunneling microscope (STM) is an extraordinary exception. It probes only the very surface of its specimen, and it does so with an atomically smooth conducting finger, held under such delicate feedback control that it follows the contours of the surface at a distance less than one atomic diameter.

Consider three images, each a showpiece from one of the scanning microscopies. Best known is the scanning electron microscope (SEM). An SEM beam can be drawn as fine as 100 atomic diameters. The now familiar images are strikingly three-dimensional, since with no need to focus the return, a beam can probe deep crevices and protrusions alike. The simple electron splash leads to return relations not unlike those we see in ordinary diffused light. One stunning full-page photograph shows a huge pile of wonderfully stacked uniform spheres. They are not atoms, of course, but tiny polystyrene spheres found in a dried film of the plastic latex, each crisply imaged ball less than half a micron across. The array bears eloquent witness to the ubiquity of close-packing, from the atomic scale of real crystals seen by the TEM, past this pinpoint of a macrocrystal, all the way up to cannonball pyramids in the park.

Scanning beams can carry not electrons but ions instead, oxygen up to indium. Heavy ions may slowly erode the specimen surface, to offer a kind of sectioning of the object over time. Collecting secondary ion products in a mass spectrometer promises the ultimate in fast microchemical analysis: secondary ion mass spectrometry. The book gives us an SIMS rendering of part of a computer chip, its intricate urban forms chemically mapped on the scale of microns, the regions of aluminum, silicon, oxides, titanium, each in its coded color. There are dozens of similar images here, a gallery of artistic micromaps, from chips to meteorites, even tissues and cells.

As matter is atomic, so images are digital. Manipulating the digital image at hand by apt numerical transformations among its digits is a familiar mainstay for aerial and orbital maps. Now it is in full sway over these micromaps. Contrast stretching, edge enhancement, grouping kindred pixels—the sample images are indeed telling, their abstract algorithms all but indifferent to imaging process or to scale. The 17th-century pioneers were justly excited by the images of fleas; their fleas now are our atoms.

The Fast Track

Rolling Thunder: A Portrait of North American Railroading, color photographs by Gary J. Benson, text by Fred W. Frailey. W. W. Norton & Company, 1991 ($49.95).

The photographer and his partner Susan Benson wait patiently in some chosen spot, "perfect clouds, beautiful light, nice super-elevated curve," for their quarry to show up in the lens. Their breath-catching natural shots show us rolling stock and railroad people pictured in their continent-wide habitats during the past few years. The text is by a knowing enthusiast, infectious in his aesthetic of bright-painted, boxy diesel-electric locomotives, canny in his informal review of the economics of U.S. and Canadian rail today, and able to name-drop the Jim Hills of our times, magnates who are somehow still obscure. Not one puff of steam (though a little black diesel exhaust) rises in these 200 striking images. Nostalgia is restricted to a long line of abandoned cabooses and a cast-off diesel or two.

Passenger rail is so modest a ripple in the flood of North American travel that the railroads seem hardly visible to the public. Yet our freight rail is healthy and modern; the tonnage it hauls is about the same as it was in 1970, now moved faster and farther and more safely on half as many cars with half as many men and women at work. Coal is still king of the carloads, followed by grain, chemicals and then motor vehicles, heavy goods in a real economy that by no means rests on gossamer information.

Most kilowatt-hours of electric power are still made by expanding steam, and most of that steam is raised by burning coal shipped by rail. Railroad track is the first segment in the long circuit that feeds your wall plugs. In a way, it is diesel oil that primes the coal-fired boilers, although the locomotives consume only 1 or 2 percent of the energy value of the coal they deliver.

A weatherworn survivor of the diesels that by the 1950s had "conquered steam" is still shoving along a push-pull commuter train near Boston's North Station. Thousands of that one rugged model were manufactured by the Electro-Motive Division of General Motors, near Chicago. That plant and the old Erie works of General Electric, renewed in the late 1970s, are the only present U.S. companies that build locomotives from scratch. New locos are pricy, up to $1.5 million for one. But a standard American-built diesel-electric delivers the horsepower of 60 new family sedans at much higher fuel efficiency, "runs weeks between shutdowns, endures climates... from 120-degree heat to minus-60-degree cold, pulls flat-out up the grades and holds back just as forcefully... going down,... keeps its footing on wet rail," if sometimes by dropping a stream of sand on the rims. It will do all that dependably for maybe 20 years. No foreign railroad "demands (or gets) so much."

Coal is a favorite subject here. You see the Wyoming fields where huge shovels strip low-sulfur coal to ship on "unit trains," dedicated long strings of big 125-ton hopper cars. The Burlington Northern takes coal out of the Powder River basin at lower cost than would a slurry pipeline, in some 40 unit trains a day, bound as far as 1,000 miles. A unit train is never sorted, "it doesn't need yards... it doesn't wait for connections... it handles easier," without stress on cars or couplings, since all its cars are either fully loaded or quite empty. "Its 10,000-ton payload dwarfs... conventional freight." It does demand state-of-the-art, billion-dollar trackage, but the investment has paid off.

The roadway and the railway can be friends. Intermodal trains have evolved over the decades since the first piggyback semitrailers on flat cars. Standard containers are now loaded on and off at trackside, for transfer to truck or

ship, by a one-operator rubber-tired lifter. The state of this art is represented by dedicated doublestackers, the containers set two layers high on special cars with lowered platforms, five cars articulated together. Many container trains roll swiftly from coast to coast, competing with the Panama Canal. Up to half the trains on big lines now are doublestacked or piggyback.

Familiar species of the tracks, like the boxcar, classification yards with yard engine and sorting hump, and the manned signal tower, are slowly declining before automation and new design. Deregulation has stimulated rationalization, productivity, energy efficiency and investors' profits, just as it has cost half the jobs. Smaller crews, well paid and independent, ride much longer stretches now. No caboose shelters a resting brakeman; only the device they call FRED is back there to monitor the rear-end brakes. A train crew is now two or three men and women, where once it was contractually five. During the 1980s, the merging railroads bought out the contractual rights to the jobs from the individual workers.

The images give the book its power. Three diesels pound up to the Tehachapi summit. Behind them bare hills blossom with strange, spreading triple wings. An aviary of rocs? No, only power windmills. Flatcars marked DODX bear a heavy piggyback load on special three-axle trucks; instantly recognizable, these are M-1 tanks rolling back to Texas from desert maneuvers. A six-locomotive train runs downhill in the foreground with its load of eastbound semitrailers, sunlit against an indigo mesa deep in shadow: "indelibly Santa Fe."

Grand Feedback

LIFE AS A GEOLOGICAL FORCE: DYNAMICS OF THE EARTH, by Peter Westbroek. W. W. Norton & Company, 1991 ($21.95; paperbound, $9.95).

Professor Westbroek leads the geobiochemistry laboratory at Leiden University. In his lab, antibodies are as much at home as geologic maps are. His brief, personal, altogether engaging book wins us by the same freshness of spirit. Can we spin the history of the earth without casting evolving life as a central actor in the long drama? That is the deep issue he addresses with clarity, candor and a "sparkle of hope."

The text (illustrated by Cees van Nieuwburg's many evocative pencil drawings) starts with a 10-square-mile piece of Dutch painter's landscape. Its dikes, meadows, bogs, flowery canals and tranquil cows border a vista of wide fields, the polders long ago reclaimed from the shallow salt sea. The author has often left his bicycle at roadside, to take coffee at the Half Moon, then to row across the lake to the 700-year-old village. Maps and cross sections of the landscape, as it is at present and as it was 1,000 years ago, show us how this complex and charming place was formed. First was a wild deltaic shoreline, shaped by riverine flow, the sea and the vegetation. Then came farm families to mine the bogs for peat and to dike and drain the flatlands. In those parts they say: God made the world, but the Dutch made Holland.

This single example opens the case for human culture as a local geologic force. Yet the world is far wider than Holland. Worldwide we find slowly drifting crustal plates that shove the seafloor beneath the mainland. The plates leave behind a seaward trench and an offshore ridge, a basin where the plate dives deep but has not yet melted, and even farther inland a volcanic, mountainous belt over the molten flow. All of these parallel the old shore.

In the epic of the plates, neither our species nor any other life-form has had much role to play. The fossils we find were bystanders. They signal waters deep or shallow or an ambience of dry land; reliable witnesses, they are hardly actors. The case here begins to seem absurd; polderland is not the world.

But life metabolizes, and its chemical turnover remains. It is that process which affixes the label of life on what might seem inanimate change. One wonderful page makes the point: crystals of calcite, or calcium carbonate, the defining ingredient of limestone, show up in two photographs. In one lie some 100 tiny rhombs of calcite, formed in a test-tube reaction. In the other are a few score tiny bits of the same mineral. But these are no mineral samples; rather they are jewels, contoured and incised plates called coccoliths. They grow from and surround a single-celled alga, *Emiliania*, an abundant, photosynthesizing organism of the plankton. Its calcareous remains lie deep in the sea, a thick layer of ooze that covers an area larger than Asia.

Limestone forms a vast portion of the solid crust, too. Did most of that rock, perhaps modified since, begin in the organic world? Abundance alone cannot decide; there are plenty of algae. Another photograph shows the creatures seen not in a microscope but from satellite orbit: a giant white bloom of *Emiliania* floats offshore, as big as Rhode Island. *Emiliania* may produce more calcite than any other organism. Once it was

easy to object that calcite would form anyway without the algae as soon as the ion concentration was high enough. All the organism did was to anticipate the event, important for its life, even for ours perhaps, but hardly for the sweep of earthly time.

The Leiden workers have shown otherwise. Coccoliths are in fact arrays of tinier single calcite crystals, articulated into a sort of microskeleton. They are laid down—the details remain far from complete—in the presence of a complex polysaccharide molecule the cell secretes. A viscous layer, found everywhere in a natural "limestone factory" such as the great lagoon behind the Florida Keys, shapes the jewel by inhibiting crystal growth along particular surfaces. Calcite formation is under biochemical control, a grand public work by the polymers we ignore as mere seafloor slime.

Now the huge tonnage of seafloor calcite matters. Heated in the lines of volcanoes, it disassociates to the greenhouse gas carbon dioxide. Afloat, its masses may produce clouds that cool the seas. What controls calcite may thus control climate and eventually build platforms out of rising seas, the atolls, to capture sunlight to fuel construction.

It will surprise no one to see that we have described a great feedback system, life nurturing itself on a geologic scale. That noble idea was put forward by James Lovelock, who coined the word "Gaia" for an earth he saw as one organism, its entire life inherently self-stabilized by evolution as much in the rocks as in the organisms.

Just how she goes about her assiduous work is unclear, for changes come from the sun, from wayward comets, from perturbed earth orbits. Like many deities, Gaia seems more our creation than our Creator. The book presents a more abstract version, named after the late geochemist Robert Garrels. This version finds an unexplained beat between the two major sulfate and carbonate geocycles, whose changes are remarkably well coupled over long times. Are they, too, linked by some unknown feedback?

Is life the true conductor of the symphony of the earth or only the clever clown who poses before the unwatching musicians to "lead" them, baton always in time but always irrelevant? Life is opportunist enough to follow change superlatively well; feedback can do no more. We do not yet know the answer, but we will await more news of *Emiliania,* enjoy this stimulating book and ponder with the author whether our species will one day turn the entire earth, like Holland, into a great "work of art," tamed, yet surely never tame.

ESSAY by Lawrence Summers

The Most Influential Investment

Educating girls quite possibly yields a higher rate of return than any other investment available in the developing world. Women's education may be unusual territory for economists, but enhancing women's contribution to development is actually as much an economic as a social issue. And economics, with its emphasis on incentives, provides guideposts that point to an explanation for why so many young girls are deprived of an education.

Parents in low-income countries fail to invest in their daughters because they do not expect them to make an economic contribution to the family: girls grow up only to marry into somebody else's family and bear children. Girls are thus less valuable than boys and are kept at home to do chores while their brothers are sent to school—the prophecy becomes self-fulfilling, trapping women in a vicious cycle of neglect.

An educated mother, on the other hand, has greater earning abilities outside the home and faces an entirely different set of choices. She is likely to have fewer, healthier children and can insist on the development of all her children, ensuring that her daughters are given a fair chance. The education of her daughters then makes it much more likely that the next generation of girls, as well as of boys, will be educated and healthy. The vicious cycle is thus transformed into a virtuous circle.

Few will dispute that educating women has great social benefits. But it has enormous economic advantages as well. Most obviously, there is the direct effect of education on the wages of female workers. Wages rise by 10 to 20 percent for each additional year of schooling. Returns of this magnitude are impressive by the standard of other available investments, but they are just the beginning. Educating women also has an impressive impact on health practices, including family planning.

Let us look at some numbers in one country as an illustration of the savings from improved hygiene and birth control. In Pakistan, educating an extra 1,000 girls an additional year would have cost approximately \$40,000 in 1990. Each year of schooling is estimated to reduce mortality of children younger than five years by up to 10 percent. Since an average woman in Pakistan has 6.6 children, it follows that providing 1,000 women with an extra year of schooling would prevent roughly 60 infant deaths. Saving 60 lives with health care interventions would cost an estimated \$48,000.

Educated women also choose to have fewer children. Econometric studies find that an extra year of schooling reduces female fertility by approximately 10 percent. Thus, a \$40,000 investment in educating 1,000 women in Pakistan would avert 660 births. A typical family-planning evaluation concludes that costs run approximately \$65 for each birth averted, or \$43,000 for 660 births.

Even beyond those savings, one can calculate that an additional year of schooling for 1,000 women will prevent the deaths of four women during childbirth. Achieving similar gains through medical interventions would cost close to \$10,000.

These estimates are of course crude. On one hand, I have failed to discount benefits to reflect the fact that female education operates with a lag. On the other, I have neglected the add-on gains as healthier, better educated mothers have not only healthier, better educated children but healthier, better educated grandchildren. (When the average mother in Pakistan has nearly 40 grandchildren, this is no small thing.)

Even with these caveats, the social improvements brought about by educating women are more than sufficient to cover its costs. Given that education also yields higher wages, it seems reasonable to conclude that the return on getting more girls into school is in excess of 20 percent, and probably much greater. In fact, it may well be the single most influential investment that can be made in the developing world.

So what can we do to promote investment in the education of girls? Scholarship funds should be established and more free books and other supplies given to girls. Providing schooling that responds to cultural and practical concerns is also essential: female enrollment depends heavily on schools' being nearby, on the provision of appropriate sanitation facilities and on the hiring of female teachers. Flexible hours and care for younger siblings can also be helpful.

Raising the primary school enrollment of girls to equal that of boys in the world's low-income countries would involve educating an extra 25 million girls every year at a total cost of approximately \$938 million. Equalizing secondary school enrollment would mean educating an additional 21 million girls at a total cost of \$1.4 billion. Eliminating educational discrimination in the low-income countries would thus cost a total of \$2.4 billion. This sum represents less than one quarter of 1 percent of the gross domestic product of the low-income countries, less than 1 percent of their investment in new capital goods and less than 10 percent of their defense spending.

When compared with investments outside the social sector, education looks even more attractive. Take power generation as an example. Projections suggest that developing countries will spend approximately \$1 trillion on power plants over the next 10 years. Because of poor maintenance and pricing problems, many of these nations use less than 50 percent of the capability of existing power plants. In a sample of 57 developing countries, the overall return on power-plant physical assets averaged less than 4 percent over the past three years and less than 6 percent over the past decade—returns that cannot even compare with those of 20 percent or more from providing education for females.

No doubt developing countries will improve their efficiency in generating power. And I have probably understated somewhat the difficulty of raising enrollment rates by neglecting capital costs and not taking explicit account of the special costs incurred in targeting girls. Nevertheless, it is hard to believe that building 19 out of every 20 planned power plants and using the savings to finance equal educational opportunity for girls would not be desirable.

There are those who say educating girls is a strategy that pays off only in the long run. This argument reminds me of a story, which John F. Kennedy used to tell, of a man asking his gardener how long it would take for a certain seed to grow into a tree. The gardener said it would take 100 years, to which the man replied, "Then plant the seed this morning. There is no time to lose."

LAWRENCE SUMMERS is chief economist at the World Bank.